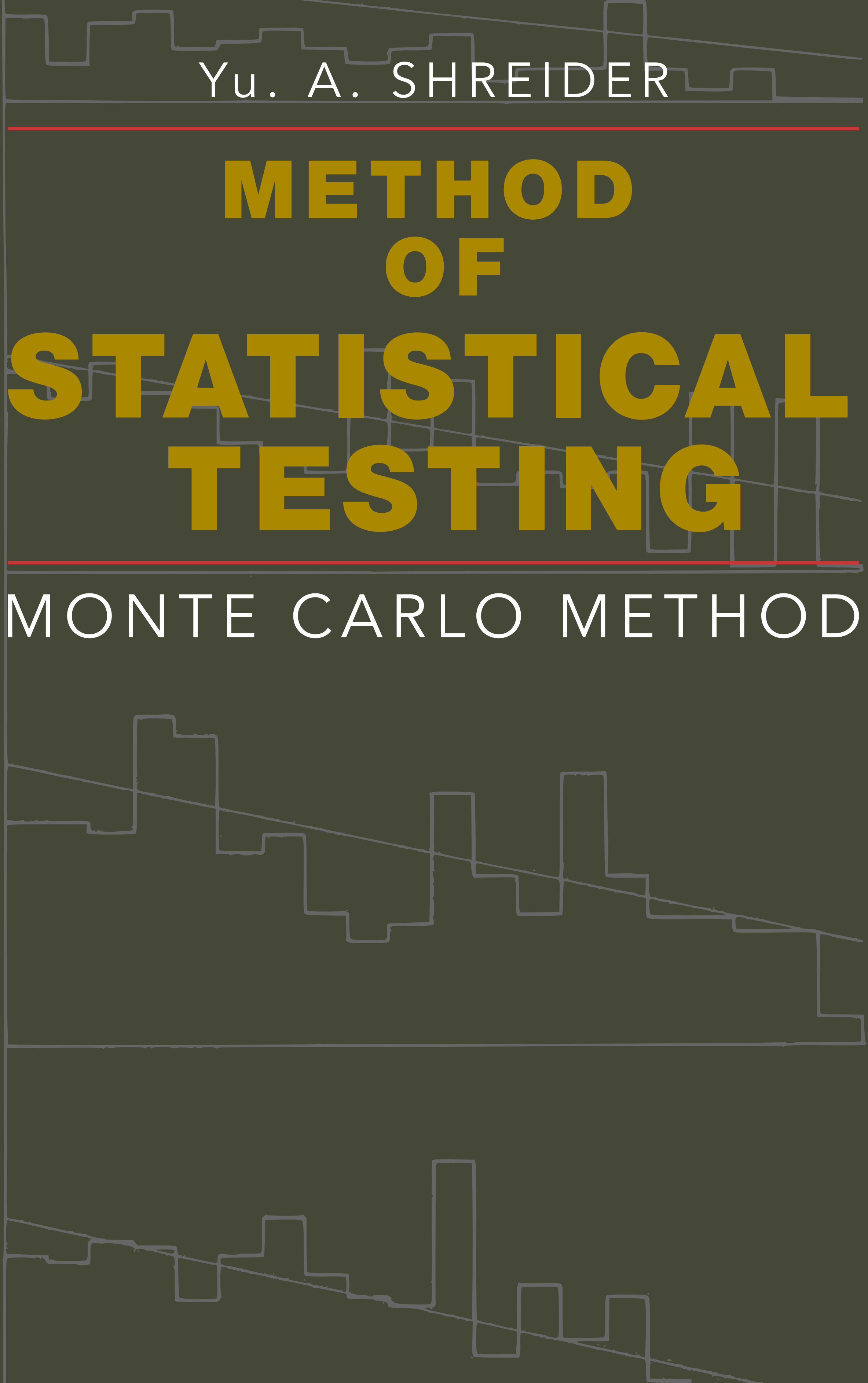
Yu. A. SHREIDER

METHOD OF STATISTICAL TESTING

MONTE CARLO METHOD



Method of Statistical Testing

MONTE CARLO METHOD

Method of Statistical Testing

MONTE CARLO METHOD

Edited by

YU. A. SHREIDER

ELSEVIER PUBLISHING COMPANY
AMSTERDAM / LONDON / NEW YORK
1964

TRANSLATED BY
SCRIPTA TECHNICA INC., WASHINGTON, D.C., U.S.A.

ELSEVIER PUBLISHING COMPANY
335 JAN VAN GALENSTRAAT, P.O. BOX 211, AMSTERDAM

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE, NEW YORK 17, N.Y.

ELSEVIER PUBLISHING COMPANY LIMITED
12B RIPPLESIDE COMMERCIAL ESTATE
RIPPLE ROAD, BARKING, ESSEX

LIBRARY OF CONGRESS CATALOG CARD NUMBER 64-15312
WITH 43 ILLUSTRATIONS AND 15 TABLES

ALL RIGHTS RESERVED
THIS BOOK OR ANY PART THEREOF MAY NOT BE REPRODUCED IN ANY FORM,
INCLUDING PHOTOSTATIC OR MICROFILM FORM,
WITHOUT WRITTEN PERMISSION FROM THE PUBLISHERS

PRINTED IN THE NETHERLANDS

Contents

PREFACE vii

CHAPTER I. Fundamentals of the Monte Carlo Method 1
1. Definition and Simple Examples of the Application of the Monte Carlo Method 1
2. Accuracy of the Monte Carlo Method and its Main Characteristics 7
3. Generation of Random Numbers 13
4. Solution of Systems of Linear Algebraic Equations . 19
5. Random Walk and the Solution of Boundary Value Problems 23
6. The Monte Carlo Method and the Simulation of Markov Processes in a Computer 31

CHAPTER II. Evaluation of Definite Integrals 39
1. Simple Applications of the Monte Carlo Method 39
2. Some Methods for Reducing the Variance 46
3. Evaluation of Multidimensional Integrals 61
4. Evaluation of Wiener Integrals 73
5. Application of Quasi-Random Points to the Monte Carlo Scheme 78

CHAPTER III. Applications of the Monte Carlo Method in Neutron Physics 84
1. The Monte Carlo Method in Elementary Particle Problems 84
2. Simple Interactions of Neutrons With Nuclei and Their Simulation 92
3. Passage of Neutrons Through a Plate 104
4. Some Methods of Calculation of the Criticality of Nuclear Reactors 118

CHAPTER IV. Application of the Monte Carlo Method to the Investigation of Mass Service or Congestion Processes, Including Queueing 127
1. General Information on Mass-Service Problems . . . 127
2. Mathematical Description of an Input Consisting of a Stream of Calls Requiring Service 129

3. Mass-Service Systems 134
4. The Generation of Random Streams of Calls 138
5. Structure of an Algorithm for Solving Mass-Service Problems by the Monte Carlo Method 148
6. Considerations On the Processing of the Results of Simulation 153

CHAPTER V. Application of the Monte Carlo Method to Information Theory 156
1. Statistical Properties of Signals and Noise 156
2. Formulation of the Basic Problems of Detection Theory 170
3. Procedure for the Solution of the Main Problems of Detection Theory 184
4. Other Problems 188

CHAPTER VI. Generating Uniformly Distributed Random Quantities by Means of Electronic Computers 196
1. Comparison of Various Methods of Generating Random Quantities 196
2. Obtaining Uniform Pseudo-Random Numbers on Computers 198
3. Criteria to Test the Quality of Uniform Pseudo-Random Numbers 208
4. Physical Generators of Uniform Random Numbers . 220
5. Tests of the Operation of Random-Number generators 238

CHAPTER VII. Transformation of Random Numbers 244
1. Properties of Quasi-Uniform Quantities 244
2. Simulation of Independent Random Events 247
3. Typical Features of the Simulation of Events by Means of Random Numbers With Few Digits 251
4. Methods of Obtaining Random Numbers With Assigned Distribution Law 252
5. Simulation of Random Vectors and Random Functions 264
6. The Simulation of Certain Multidimensional Quantities 266

APPENDICES 273

REFERENCES 281

INDEX 299

Preface

The present volume of the Mathematical Reference Library is devoted to the so-called method of statistical trials (the Monte Carlo method). In contrast with the earlier volumes, which were devoted to the classical divisions of mathematics and a sharply delineated subject matter with well-established terminology and traditions of exposition, the mathematical methods examined in the present volume have been developed only in the last thirteen years.*

These methods, which are applied in the most varied fields of computational mathematics, are unified by a single common idea.

They are based on the principle of simulating a statistical experiment by computational techniques and recording the numerical characteristics obtained from this experiment. Therefore, all these methods are united under the common name of statistical trials or the Monte Carlo method. The solution of numerical problems by this method is closer in spirit to physical experiments than to classical computational methods. Error in the Monte Carlo method cannot be sufficiently well evaluated in advance and, as a rule, is found by determining the mean squares for the simulated quantities. In a number of cases the solution cannot be accurately reproduced. The solution is stable with respect to single errors in operation of the given electronic computer.

It is the purpose of the present volume to show the fundamental distinctive features of the Monte Carlo method, giving a sufficiently thorough discussion of the facilities and typical procedures employed and the principal regions of application.

In Chapter I (by Yu. A. Shreider) the distinctive features of the method are expounded and typical examples of its application in problems of purely mathematical computation are examined.

---

*Essentially, application of the Monte Carlo method dates back to the work of N. Metropolis and S. Ulam [203], which appeared in 1949. The term "Monte Carlo" appeared first in this work.

In Chapter II (by I. M. Sobol') the principal region of application of the method—the evaluation of multidimensional integrals—is examined in great detail. In it, many methods of statistical simulation of integrals are analyzed and the accuracy of the calculations is investigated.

Chapter III (by I. M. Sobol') and Chapter V (by V. G. Sragovich) are devoted to applications of the Monte Carlo method in those regions of physics and technology in which it is most widely and successfully used. (The method was first applied in the field of neutron physics.) These chapters discuss random processes (the passage of a beam of neutrons through a scattering medium, fluctuations in a radio signal, etc.) that are so complicated that analytic description of them is practically impossible. However, successful investigation of such processes can be achieved by their statistical simulation in digital computers. These processes have special peculiarities and in their simulation many special methods have been devised which it was found convenient to put into separate chapters.

Chapter IV (by N. P. Buslenko) discusses application of the Monte Carlo method to investigation of mass-service processes (including queueing). This field, which has been developed only recently, is related to simulation of the most complicated systems of equations and to operations research. It holds great promise for use in mathematical cybernetics.

In Chapters VI (by D. I. Golenko) and VII (Sections 1, 2, 3, and 5 by N. P. Buslenko, Section 6 by V. G. Sragovich, and Section 4 by N. P. Buslenko and V. G. Sragovich together) the methods of organizing statistical experiments for general-purpose digital computers are examined. Important questions on every application of the method of statistical trials are examined in these chapters.

Among the important fields of application of the Monte Carlo method, two remain unelucidated: self-organizing systems and reliability of complicated radio-electronic assemblies (e.g., computers). However, each of these fields is still in its infancy and the inclusion of appropriate material in a reference work such as this would be problematical. Questions relating to the first of these fields are partially explained in the work by R. Bush and F. Mosteller, Stochastic Models for Learning (John Wiley and Sons, 1955).

Since the bibliography for the separate chapters coincides in a number of cases and also for reasons of convenience, a single bibliography on the Monte Carlo method appears at the end of the book (compiled by I. M. Sobol'). All citations in the text refer to this bibliography.

General problems of the Monte Carlo method have been explained in a recent book by N. P. Buslenko and Yu. A. Schreider [11]. The present volume differs from it in its more elementary and detailed exposition of the material and in its references to a wider class of applications.

The text does not discuss structural aspects of digital computers designed for solving problems by the Monte Carlo method and, in particular, questions concerning the design of special-purpose digital computers.

A large group of authors participated in writing this volume. Although the work was done on a chapter basis, some unavoidable overlap has remained in the content, and there is also some variation in the style of exposition.

The work is intended for a wide circle of readers varying from persons seeking familiarity with the fundamentals of application of the method to persons who are interested in comparatively detailed aspects of the simulation of physical processes. With the exception of Chapters III and V, which are designed for readers familiar with certain general information on neutron physics and radio engineering, understanding of the greater part of the material presented in this book requires of the reader only mathematical preparation in a university course in mathematical analysis and also a knowledge of the fundamentals of probability theory, since probability concepts run like a fine thread throughout the entire book. It is assumed that the reader is familiar with the fundamentals of random events and random variables and with their probability characteristics (the probability of an occurrence, mathematical expectation, variance). Furthermore, he must have a grasp of the concept of normal distribution, Lyapunov's theorem and, for certain chapters, of the elements of mathematical statistics. An understanding of Markov processes would also be very desirable.

The authors hope that this reference book will prove useful to people who deal with the Monte Carlo method and its application and, what is no less important, will give many readers an understanding of the usefulness of the Monte Carlo method in the solution of various problems. The authors express their gratitude to V. D. Rozenknop for his valuable advice during preparation of this work.

Yu. Shreider

Chapter I

Fundamentals of the Monte Carlo Method

1. DEFINITION AND SIMPLE EXAMPLES OF THE APPLICATION OF THE MONTE CARLO METHOD

The Monte Carlo method is a method of solving various problems in computational mathematics by constructing for each problem a random process with parameters equal to the required quantities of that problem. The unknowns are determined approximately by carrying out observations on the random process, and by computing its statistical characteristics which are approximately equal to the required parameters.

For example, the required quantity x may be the mathematical expectation $\mathsf{M}\xi$ of a random quantity. The Monte Carlo method for the approximate determination of x will then involve the N-fold sampling of ξ, yielding a series of independent values $\xi_1, \xi_2, \ldots, \xi_N$, and the calculation of the mean

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \ldots + \xi_N}{N}.$$

In accordance with the law of large numbers, when N is sufficiently large the mean is given by

$$\bar{\xi} \approx \mathsf{M}\xi = x,$$

which holds with probability approaching unity. It follows that the quantity $\bar{\xi}$, which is deduced from observations on the random process, is approximately equal to the required quantity x.

Consider a simple example. Suppose it is necessary to calculate the probability w that the total number of "hits" scored with an arrow shot at a target is even after ten attempts. If the probability of scoring a hit as a result of a single attempt is p, then the required probability w can be calculated from the formula

$$w = \sum_{k=0}^{5} C_{2k}^{10} p^{2k} (1 - p)^{10-2k}. \qquad (1.1)$$

This formula is clearly correct, since the general term following the summation sign is the probability that the number of hits is exactly equal to $2k$. The calculation of the probability w using equation (1.1) and tabulated values of C_{2k}^{10} will involve 18 multiplications and 6 additions.

Alternatively, one could carry out N series of 10 shots each and determine the number of cases, L, where a series of ten shots resulted in an even number of hits. When N is sufficiently large, the quotient L/N will be a good approximation to the required probability w. It will be shown below that in order to obtain an estimate of w which will be correct to two decimal places, it is necessary to carry out 10,000 series of trials each involving 10 shots.

In the example just quoted, it is simpler to calculate the probability from equation (1.1) than to make 100,000 shots.

This example is an illustration of how an unknown quantity may be determined from a real experiment. More precisely, the Monte Carlo method is usually understood to involve the construction of an artificial random process exhibiting all the required properties, but realized by ordinary computation means, e.g., pencil, paper, numerical tables, calculating machines, and, occasionally, simple means of generating random quantities (the so-called random-number generators; cf. Chapter VI).

In practice, the Monte Carlo method has found wide application only in the course of the exploitation of large computers. We shall see later that the method is particularly convenient with digital computers. As a rule, the Monte Carlo method is used in general-purpose computers, although it is frequently feasible to construct special-purpose machines for particular classes of problems.

We shall now consider how the above probability w can be determined with the aid of simple computational means. Instead of the arrow we shall set up a process which will involve the use of a spinning top and a clock with a second hand, and also pencil and paper. We shall assume for simplicity that $p=\frac{1}{5}$. Instead of a series of 10 shots we shall spin the top 10 times, and when it falls on the floor we shall look at the clock and read off the position of the second hand. If the second hand indicates a value τ in the range $0 \leqslant \tau < 12$ then we shall score a "hit." Again, if in a series of 10 spins we observe an even number of hits, then the series will be regarded as "successful." Suppose that N series of such experiments, each involving 10 spins, result in L successes. The random quantity L/N will then be distributed in precisely the same way as the analogous quantity obtained above with a real arrow, and will therefore be approximately equal to the required probability w.

The experiment involving the spinning top is more economical but is no less time-consuming than the "real" experiment. In order to accelerate the process it is necessary to use an electronic computer. The random process is then simulated with the aid of the computer, and its statistical characteristics are evaluated and printed out by the machine in the form of a final solution.

In particular, the above example may be realized on a computer as follows. Many computers incorporate random-number generators (cf. Chapter VI) which produce in each cycle the value ξ of a random quantity, which is uniformly distributed in the range $(0, 1)$. Instead of shooting an arrow or spinning a top, one can note the value ξ produced by the generator, and check whether the inequality $\xi < p$ is satisfied. If it is, then we shall consider that a hit has been scored. It is easy to see that the probability for the inequality $\xi < p$ to be satisfied, which is equal to the probability of scoring a hit, is, in fact, p.

Let us select a series of 10 such values of ξ. If this series includes an even number of "hits" then it will be regarded as a "success." Suppose that among N series there are L successes. The quantity L/N can then be interpreted exactly as above.

The simulation of a single shot by this method involves two operations, so that the imitation of the entire process with $N = 10{,}000$ repeated series will involve 200,000 operations. In the case of the "Strela" computer, this would require 100,000 seconds of time, which is already much faster than the actual experiment of 100,000 shots, but is still much higher than the 24 operations which are necessary if the probability is to be computed from equation (1.1).

For problems involving the determination of the accuracy of artillery fire from a large number of guns, the simulation of the process with a computer will occupy much less time than the analytical calculation of the required quantities with the aid of the same computer. Here, the method of statistical trials is virtually the only acceptable method of obtaining the final result.

Occasionally, the analytical formulation of the problem (for example, the boundary value problem for the Laplace equation) forms the starting point. The appropriate random process (for example, the random walk process described in Section 5 of this chapter) is then found, and a study is made of this process. In other cases, the starting point is a random process whose analytical description is of little use in practice, or is not considered at all. An illustration of this is the above calculation of the probability of hitting a target under given firing conditions, or the determination of the parameters of a mass-servicing process.

In the above example we were concerned with the precise simulation of a random process. In practice it is more usual to compare the process under investigation with a simplified artificial process which can be simulated in a computer and is in some sense an approximation to the original process. Thus, when the Monte Carlo method is used to solve the boundary value problem for the Laplace equation in a region G, the corresponding random process is the Brownian motion in the region G with continuous time and precipitation on the boundary of the region. On the other hand, the situation is usually simulated by a random walk, with a discrete time, over a lattice inscribed in the region G, so that at each instant there is a "jump" of the Brownian particle from a site of the lattice to a neighboring site (cf. Section 5). This simplification is usually dictated both by the absence of complete information about the real process and by computer limitations which prevent the simulation of a complicated process in an acceptable length of time.

The Monte Carlo method has been most successful in those fields where the basic mathematical problem involves the investigation of some random process. Thus, problems in neutron physics are formulated in a probabilistic way, the separation of signals from a random-noise background is a probabilistic problem, and so on. However, there are computational problems whose formulation is not related to the theory of probability, but which can be successfully solved by the Monte Carlo method. The most typical examples are the boundary value problems for elliptical equations (for example, the Laplace equation), and the related problems for parabolic equations (for example, the heat transfer equation).

These equations are closely connected with certain random diffusion-type processes. It follows that the solution of these equations can be conveniently carried out by simulating such processes. A method whereby this can be done is discussed in Section 5. It is interesting to note that when the relation between the boundary value problems and random processes was first discovered, the main interest was focused on the fact that the methods of the theory of differential equations could be used to study a wide class of random processes. In the Monte Carlo method, this classical situation is reversed: the simulation of random processes is found to be a very convenient method of finding the solutions of differential equations.

A second example is the solution of linear algebraic equations. Section 4 gives an example of a statistical simulation of a process whose characteristics are used to find the unknowns in a system of linear algebraic equations of the form $x = Ax + b$, where A is a matrix which is approximately equal to the unit matrix. Methods which may be used for more general systems have been discussed by Shreider [68].

The development of methods based on probability theory for evaluating integrals is of particular importance. Since a probability can always be regarded as a measure, it follows that the determination of the probabilities of some given events, or their mathematical expectations, can always be reduced to the evaluation of an integral. Consider the integral

$$\int_0^1 f(x)\,dx,$$

where $f(x)$ is such that $0 \leqslant f(x) \leqslant 1$ for $0 \leqslant x \leqslant 1$, i.e., the values of $f(x)$ lie between 0 and 1. It is required to find the area S of the region G bounded by the curve $y = f(x)$, the x-axis and the ordinates $x = 0$, $x = 1$ (Fig. 1). It should be noted that the limitation imposed on the function $f(x)$ is unimportant because the scale can, clearly, be altered. Suppose that we choose a random point in the square $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ by taking one whose coordinates are independently uniformly distributed in the interval (0, 1). What is the probability p that the point will be in the region under the curve? Suppose that one such point is (ξ, η). This point will lie in the square, since $0 \leqslant \xi \leqslant 1, 0 \leqslant \eta \leqslant 1$. It is clear that the probability p is equal to the area S, i.e., the required integral.

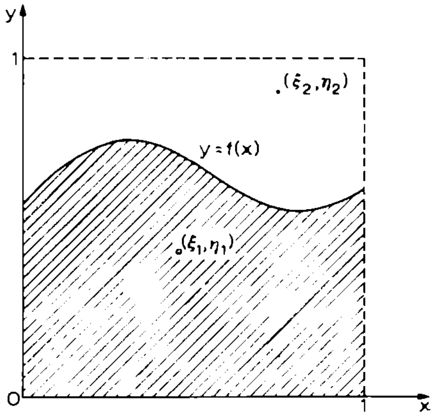


Fig. 1.

Suppose we have some way of obtaining independent and uniformly distributed quantities ξ and η. We will take the first pair of these quantities and check whether the condition

$$f(\xi) > \eta \tag{1.2}$$

is fulfilled. If it is, then the chosen point (ξ, η) does lie in the region G under the curve. Next, let us take N pairs of random quantities, and for all of them verify whether the inequality (1.2) is satisfied. Suppose that for L pairs of these N selections the inequality (1.2) is satisfied. The quantity L/N will then be approximately equal to the probability that the random point lies in the region G, i.e.,

$$\frac{L}{N} \approx p = \int_0^1 f(x)\,dx.$$

The estimated error in the integral evaluated in this way will again depend on the number of trials N.

In both the examples discussed above, the simulation of the process involved the estimation of the probability of some event A. A more general case is one in which an estimate is made of the unknown mathematical expectation of a random quantity η. A simple example will involve the evaluation of the integral $\int_0^1 f(x)\,dx$. Suppose ξ is a quantity which is uniformly distributed in the interval $(0, 1)$. The quantity $\eta = f(\xi)$ will then have the mathematical expectation

$$a = \mathsf{M}\eta = \int_0^1 f(x)\,dx.$$

Thus, in order to evaluate the integral, it is necessary to choose N independent values $\xi_1, \xi_2, \ldots, \xi_N$ of the quantity ξ and form the arithmetic mean

$$\bar{\eta} = \frac{f(\xi_1) + f(\xi_2) + \ldots + f(\xi_N)}{N}.$$

The quantity $\bar{\eta}$ will be an approximation to the integral. We note that an error in a few values of $f(\xi_i)$ will be smoothed out when N is large, i.e., the method is independent of random machine faults.

The examples considered in this section will serve as an illustration of the general characteristics of the Monte Carlo method, i.e., the necessity of carrying out extensive series of calculations of a given type, and the smoothing out of errors. In what follows, other characteristics of the method will become apparent, for example, the use of relatively small memory stores for intermediate results, and adaptability to multidimensional problems. All these characteristics show that the Monte Carlo method is very suitable for calculations on large digital computers.

2. ACCURACY OF THE MONTE CARLO METHOD AND ITS MAIN CHARACTERISTICS

Let us now consider the accuracy of the Monte Carlo method. Suppose that an event A which occurs with probability p is to be simulated. Let ξ_i be a quantity which is equal to unity if the ith trial results in A, and zero if the event A does not occur. Thus, the total number of trials for which the event A does occur is

$$L = \sum_{i=1}^{N} \xi_i,$$

where N is the total number of trials. Individual trials will be assumed to be independent.

The relative frequency of occurrence of the event A is equal to L/N and is a random quantity whose mathematical expectation is

$$\mathsf{M}\left(\frac{L}{N}\right) = \frac{1}{N}\mathsf{M}(L) = \frac{1}{N}\sum_{i=1}^{N}\mathsf{M}\xi_i = \frac{Np}{N} = p$$

and whose dispersion is

$$\mathsf{D}\left(\frac{L}{N}\right) = \frac{1}{N^2}\mathsf{D}(L) = \frac{1}{N^2}\sum_{i=1}^{N}\mathsf{D}\xi_i = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N}.$$

In the above expressions we have used the fact that the mathematical expectation for each of the ξ_i is

$$\mathsf{M}\xi_i = 0(1-p) + 1\cdot p = p,$$

and the dispersion is

$$\mathsf{D}\xi_i = \mathsf{M}\xi_i^2 - (\mathsf{M}\xi_i)^2 = 0(1-p) + 1\cdot p - p^2 = p - p^2 = p(1-p).$$

According to the law of large numbers (Bernoulli's theorem), the relative frequency of occurrence of the event A, which is equal to L/N, is approximately equal to the probability p. More precisely, for any $\varepsilon > 0$ and any $\delta > 0$ there exists a number of trials N which is such that with a probability greater than $1-\varepsilon$, the relative frequency of occurrence of the event A will differ from the probability p, that the event A will occur, by an amount less than δ:

$$\left|\frac{L}{N} - p\right| < \delta. \tag{1.3}$$

Since p is the required quantity, while $\frac{L}{N}$ is its approximate value obtained by the Monte Carlo method, it follows that the difference $\frac{L}{N} - p$ is the error of the Monte Carlo method.

It is clear from the above that this error can only be estimated probabilistically with a degree of reliability $1 - \varepsilon$. In what follows we shall usually assume that the degree of reliability is 0.99 or 0.997.

It should be noted that the use of computers does, in one sense, remove the distinction between the Monte Carlo method and ordinary computational methods. In fact, whenever a complicated computational problem is solved with the aid of a computer, it is always necessary to consider the possibility of random error (due to a fault or a rounding-off procedure) and therefore the final result can only be regarded as reliable with a certain probability approaching unity. Moreover, in complicated computational problems, the error can only be estimated from the results of the calculation.

The left-hand side of the inequality (1.3) can always be estimated with the aid of the Chebyshev inequality,

$$\left|\frac{L}{N} - p\right| \leqslant \sqrt{\frac{p(1-p)}{\varepsilon N}}, \tag{1.4}$$

where ε is, as above, the probability that the estimate (1.4) does not hold. This estimate is derived in textbooks on the theory of probability. The most important fact is that the estimate is of the order of $N^{-\frac{1}{2}}$. In other words, the error δ involved in the value given by the Monte Carlo method for the probability that the event A will occur is given by

$$\delta \sim \frac{1}{\sqrt{N}}. \tag{1.5}$$

Although the estimate (1.4) may be considerably improved, the relation given by (1.5) will always hold. It has a very important bearing on the accuracy of the Monte Carlo method and the limits of its applicability.

It is clear from (1.5) that the error δ in the approximate solution of a problem by the Monte Carlo method can be reduced by increasing the number of trials N, i.e., by increasing the computational time. For example, the time necessary to complete the solution must be increased by a factor of 100 if the accuracy is to be increased by one order of magnitude. It follows that the Monte

Carlo method cannot yield highly accurate solutions. In practical problems the Monte Carlo method will yield an accuracy of the order of 0.01-0.001 of the maximum value, unless special methods for accelerating the calculations are employed.

It will be shown later that the Monte Carlo method is very suitable for the solution of multidimensional problems. These problems do not usually require high accuracy, and hence the above disadvantage of the method is not as important as it might appear at first sight.

Moreover, it should be noted that when the probabilities are computed by the Monte Carlo method (for example, the probability of hitting a target, the probability that a neutron will penetrate through the shielding of a reactor, and so on), high accuracy is frequently unnecessary in view of the very nature of the problem. Thus, it is immaterial in practice whether a target will be hit with a probability of 0.901 or 0.902.

Suppose it is necessary to simulate a process for which each ith trial of N independent trials yields a quantity ξ_i. Suppose that this quantity has a finite mathematical expectation $\mathsf{M}\xi_i = a$ and a variance $\mathsf{D}\xi_i = \sigma^2$. The arithmetic mean

$$\bar{\xi} = \frac{L}{N} = \frac{\sum_{i=1}^{N} \xi_i}{N}$$

is then an approximate value of the required mathematical expectation a. The quantity $\bar{\xi}$ is the result of the solution of the problem by the Monte Carlo method. The deviation of this quantity from the required mathematical expectation is in fact the error of the method.

As above, the estimate of the error is associated with the probability $1-\varepsilon$. This means that the estimate

$$\delta = |\bar{\xi} - a| \leqslant \sigma \sqrt{\frac{1}{\varepsilon N}}, \tag{1.6}$$

which can be obtained in the same way as (1.4) from the Chebyshev inequality, holds with a probability not less than $1-\varepsilon$. The relation

$$\delta \sim \frac{1}{\sqrt{N}}$$

will hold as before with all the resulting consequences.

Equation (1.6) clearly indicates the significance not only of the number of trials, but also of the variance of ξ_i. One should always bear in mind the possibility of reducing this variance, using, when necessary, special methods which improve the convergence. In fact, the number of trials N may be reduced by reducing σ in the estimate (1.6), and this will, of course, also reduce the time necessary for the solution of the problem. Let us now improve the estimate (1.6), and as a special case derive (1.4). To do this, we note that the quantity $L=\xi_1+\xi_2+\ldots+\xi_N$ is the sum of a large number of random quantities. This is also true of the quantity

$$\bar{\xi}=\frac{\xi_1}{N}+\frac{\xi_2}{N}+\ldots+\frac{\xi_N}{N}. \tag{1.7}$$

It follows that the distribution of $\bar{\xi}$ may be obtained from the limit theorems of the theory of probability. Let r_α be a quantity which is such that with a probability α we have

$$|\bar{\xi}-a|\leqslant r_\alpha.$$

The quantity r_α may serve as an estimate of the error, which in general is better than (1.6).

The most important case is that in which the quantity $\bar{\xi}$ is distributed in accordance with the normal law (Gaussian distribution). The fact that $\bar{\xi}$ is distributed very nearly in this way can be established from some very general properties of the quantities ξ_i. This fact is ensured, for example, if the quantity

$$\frac{b}{\sigma^3\sqrt{N}} \tag{1.8}$$

is sufficiently small, where b is the so-called third moment of ξ_i. Since the values of N which are commonly employed in the Monte Carlo method are of the order of 10^3-10^5, it follows that the Gaussian distribution will as a rule be obeyed. The only exception will be the simulation of events occurring with a low probability p, in which case the Poisson distribution will hold for integral values of L. The condition that the limit law for L is the Poisson distribution is

$$p\sim\frac{1}{N}. \tag{1.9}$$

In other words, it is necessary that the quantity pN should not be too large. If this condition is satisfied, the quantity given by (1.8) will assume the form

$$\frac{b}{\sigma^3\sqrt{N}}=\frac{p(1-p)(1-2p)}{p^{3/2}(1-p)^{3/2}\sqrt{N}}=\frac{(1-2p)}{(1-p)^{1/2}\sqrt{pN}},$$

which, in view of (1.9), is not of the order of zero, i.e., contradicts the normality condition.

Consider now the usual case, in which $\bar{\xi}$ is distributed very nearly in accordance with the Gaussian law. When the error is known with probability $\alpha = 0.997$ we have $r_\alpha = 3\sigma_0$, where σ_0 is the standard deviation of $\bar{\xi}$ (this is the so-called three-sigma rule).

The variance σ_0^2 of the arithmetic mean $\bar{\xi}$ is related to the variance σ^2 of ξ_i by the formula

$$\sigma_0 = \frac{\sigma}{\sqrt{N}}, \tag{1.10}$$

and hence the error of the Monte Carlo method may be estimated from

$$\delta = |\bar{\xi} - a| \leqslant \frac{3\sigma}{\sqrt{N}}. \tag{1.11}$$

This estimate is exact. If the quantity on the right-hand side of (1.11) is reduced, then the corresponding probability is reduced. The expression (1.11) shows that the accuracy of the Monte Carlo method depends only on the number of independent trials and the variance. In particular, the number of operations involved in a transition from one-dimensional integrals to multidimensional integrals will increase only as a result of the increase in the complexity of the integrand and in the number of uniformly distributed quantities employed, i.e., in practice it will be directly proportional to the dimensionality. At the same time, the number of operations in the case of ordinary quadratures increases as K^n where n is the multiplicity of the integral. This shows that the Monte Carlo method is very suitable for multidimensional problems. This will be seen from examples which will be considered later.

Equation (1.11) can also be interpreted in another way. Let δ be the required accuracy of the calculations, and ν the average number of computer operations per trial. The total number of operations F will then be given by

$$F \approx N\nu \approx \frac{9\nu\sigma^2}{\delta^2}, \tag{1.12}$$

and the total time for the solution of the problem by

$$T \approx N\nu\tau \approx \frac{9\nu\sigma^2\tau}{\delta^2}, \tag{1.13}$$

where τ is the computer time per operation. It is clear from this formula that the time T must be increased by two orders of magnitude if the accuracy is to be improved by one order of magnitude.

It is usually relatively difficult to estimate the variance σ^2 in advance (i.e., prior to the solution of the problem). Theoretical estimates are, as a rule, much too high. In solving specific problems, the error may be estimated by substituting the statistical estimate of the variance

$$\Delta=\frac{(\xi_1-\bar{\xi})^2+(\xi_2-\bar{\xi})^2+\ldots+(\xi_N-\bar{\xi})^2}{N-1}$$

into the right-hand side of (1.11) instead of the theoretical variance σ^2. The statistical estimate is obtained as a result of the simulation of the values of the random quantity.

Thus, the accuracy which may be achieved with the Monte Carlo method can only be properly estimated in the process of solution. This fact is analogous to the well-known result that the accuracy of a physical experiment can be reliably determined only from the results of the experiment itself.

The estimates given above were concerned with absolute errors. Let us return now to the simulation of an event A occurring with probability p, and estimate the relative error in the determination of p with the aid of the frequency L/N. In this case the estimate (1.11) will be of the form

$$\delta=\left|\frac{L}{N}-p\right|\leqslant 3\sqrt{\frac{p(1-p)}{N}},$$

and hence the estimate for the relative error is

$$\frac{\delta}{p}\leqslant 3\sqrt{\frac{1-p}{pN}}. \qquad (1.14)$$

Thus, the required number of trials can be determined in terms of the admissible relative error d from the formula

$$N\approx\frac{9(1-p)}{pd^2}, \qquad (1.15)$$

which shows that the number of trials is inversely proportional to the required probability p.

This means that very small probabilities cannot be determined by the Monte Carlo method. When the magnitude of N as given by (1.15), or, more precisely, the time $T=N\nu\tau$ for the solution of the problem, is inadmissibly large, then it is necessary to try to transform the problem into another with reasonable values of the probability p.

We note that in deriving (1.14), use was made of an estimate based on the assumption that the distribution law is normal, although we were concerned with a low probability p. The point is that when $pN \sim 1$, it is impossible in principle to obtain a low relative error, and it is therefore necessary to demand that the condition $pN \gg 1$ should be satisfied. This in itself leads to a distribution law for the frequency M/N which approaches the normal distribution.

3. GENERATION OF RANDOM NUMBERS

In order to be able to solve practical problems by the Monte Carlo method, it is necessary to have a random-number generator with a sufficiently varied store of distribution laws.

In this section we shall consider various methods of generating random events and random numbers. These problems will be considered in greater detail in Chapters VI and VII.

It turns out that uniformly distributed random quantities are the most important. They may be used to simulate random events and random quantities obeying various distribution laws.

Suppose we have a random quantity ξ which is uniformly distributed in the interval (0, 1). Let us consider the problem as to how this random quantity may be used to simulate an event A which occurs with probability p. Let us define A as the event that "the chosen ξ satisfies the inequality $\xi < p$." It is easy to see that the probability that the event A will occur is

$$\int_0^p dx = p,$$

and hence the event A will occur with the given probability p.

The uniformly distributed random quantities ξ may be used to obtain other quantities which obey practically any given distribution law $F(x)$. All that is necessary is to replace ξ by some monotonic function of ξ, for example, $\eta = f(\xi)$. The quantity η will follow the distribution law

$$F(x) = \int_{-\infty}^{f^{-1}(x)} h(t)\,dt, \tag{1.16}$$

where $f^{-1}(x)$ is the function which is inverse of $f(x)$ and $h(x)$ equals unity for $0 < x < 1$, and zero outside the interval (0, 1).

In order to obtain the exponential distribution law for which

$$F(x)=\begin{cases} 0 & \text{for} \quad x \leqslant 0, \\ 1-e^{-\omega x} & \text{for} \quad x > 0, \end{cases}$$

it is sufficient to take

$$f(x) = -\frac{1}{\omega}\ln(1-x). \tag{1.17}$$

In fact, the function which is inverse to $f(x)$ will be of the form $f^{-1}(x) = 1 - \exp(-\omega x)$, and in accordance with (1.16)

$$F(x) = \int_{-\infty}^{1-e^{-\omega x}} h(t)\,dt = \begin{cases} 0 & \text{for} \quad x \leqslant 0, \\ 1-e^{-\omega x} & \text{for} \quad x > 0, \end{cases}$$

Since for $x > 0$ the upper limit lies inside the interval (0, 1), the quantity η will in this case be distributed in accordance with the exponential law.

In order to obtain quantities distributed in accordance with the Gaussian law, one can use yet another construction. According to Lyapunov's theorem, the quantity

$$\xi = \sqrt{\frac{12}{N}}\left\{\xi_1 + \xi_2 + \dots + \xi_n - \frac{n}{2}\right\}$$

will approximately follow the normalized Gaussian law for sufficiently large n, where $\xi_1, \xi_2, \dots, \xi_n$ are independent, uniformly distributed quantities. In practice it is quite sufficient to take $n = 5$. The values of the corresponding empirical distribution obtained by this device with the aid of the "Ural" computer are given in Table 1. In this calculation 1,000 values of ξ were taken to compute the empirical distribution law.

Thus, the main problem consists in finding a method of obtaining uniformly distributed random quantities.

Consider the discrete random quantity x_i which can assume the values 1 or 0, each with probability 0.5. For example, we could toss a coin and define $x_i = 1$ as corresponding to "heads" and $x_i = 0$ as corresponding to "tails."

Consider now the infinite sequence $x_1, x_2, \dots, x_i, \dots$ and look upon this series of zeros and units as the binary digits of some number ξ given by

$$\xi = \sum x_i 2^{-i}.$$

The number ξ is a random number lying in the range $0 < \xi \leqslant 1$. The probability that ξ will lie in the interval (0, 0.5) is equal to 0.5,

Table 1

x	Empirical law	Normal law	Deviation
−2.7	0.003	0.002	0.001
−2.4	0.012	0.007	0.005
−2.1	0.023	0.017	0.006
−1.8	0.045	0.035	0.010
−1.5	0.068	0.066	0.002
−1.2	0.114	0.114	0.000
−0.2	0.183	0.183	0.000
−0.6	0.279	0.273	0.006
−0.3	0.396	0.382	0.014
0.0	0.503	0.500	0.003
0.3	0.611	0.618	−0.007
0.6	0.720	0.727	−0.007
0.9	0.815	0.817	−0.002
1.2	0.879	0.886	−0.007
1.5	0.932	0.934	−0.002
1.8	0.970	0.965	0.005
2.1	0.988	0.983	0.005
2.4	0.994	0.993	0.001
2.7	0.999	0.998	0.001

the probability that it will lie in the interval (0, 0.25) is equal to 0.25, and so on. In general, the probability that ξ will lie in any interval of the form $\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right)$ is equal to the length of the interval $\frac{1}{2^n}$. It follows that ξ is a uniformly distributed random quantity. Hence, we have a method of simulating a uniformly distributed random quantity. The method requires the availability of an infinite sequence of independent random quantities $\{x_i\}$ which are looked upon as the binary digits of a number ξ..

In practice, it will be necessary to cut off this sequence at a finite value of n and obtain a quantity ξ^* whose distribution has the form of a step-function $F_n(x)$. This function will approximate the uniform distribution law given by the function $F(x)$ which is defined by

$$F(x)=\begin{cases} 0 & \text{for} \quad x<0, \\ x & \text{for} \quad 0\leqslant x\leqslant 1. \\ 1 & \text{for} \quad x\geqslant 1. \end{cases}$$

The deviation can be estimated from the inequality

$$|F(x)-F_n(x)|<\frac{1}{2^n}.$$

There are two main types of physical generators for the random numbers x_i. The first type makes use of the emission from radioactive matter, and the second is based on the intrinsic noise of electronic tubes.

Consider the first type and suppose that we have a radioactive source. A counter is used to count the number of radioactive particles which are emitted in a time Δt. If the number of these particles is odd then x_i is taken to be unity. If the number is even, then x_i is taken to be zero. It is usually considered that the probability w_k that the counter will record k particles in the time Δt is given by the Poisson distribution

$$w_k=\frac{(\lambda\,\Delta t)^k}{k!}\,e^{-\lambda\,\Delta t}.$$

The probability that an even number of particles will be recorded is then given by

$$p_0=\sum_{k=0}^{\infty} w_{2k}=\sum_{k=0}^{\infty}\frac{(\lambda\,\Delta t)^{2k}}{(2k)!}\,e^{-\lambda\,\Delta t}. \tag{1.18}$$

Substituting $z=\lambda\,\Delta t$ we have

$$p_0=\sum_{k=0}^{\infty}\frac{z^{2k}}{(2k)!}\,e^{-z}=e^{-z}\cosh z=\frac{1+e^{-2z}}{2}.$$

Hence, it is clear that the probability of recording an even number of particles in the time Δt is given by

$$p_0=\frac{1+e^{-2\lambda\,\Delta t}}{2}.$$

When $\lambda\,\Delta t$ is sufficiently large, p_0 will tend to 0.5.

The quantity $\lambda\,\Delta t$ is equal to the mathematical expectation of the number of particles recorded in Δt. Hence, it is clear that the time Δt necessary to obtain a single binary digit must be such that the average number of particles recorded in Δt is sufficiently large. When $\exp(-2\lambda\,\Delta t)=0.01$, then $\Delta t\sim\frac{1}{2\lambda}\ln 100$. Hence, $\lambda\,\Delta t\sim\ln 10$, which corresponds to roughly three particles.

When it is required to obtain m binary digits, then the time will be greater by a factor of m. Assuming that with the accuracy usually required the quantity ξ is taken to 15 binary digits, it follows that the time necessary to obtain a single uniformly distributed number is such that the total number of particles recorded in that time by the counter is 40 to 50. When it is necessary to obtain normally distributed numbers, then this time will increase accordingly. This constitutes a limitation of the applicability of this method of generating random numbers.

Consider now the second method. Electronic tubes are always found to exhibit intrinsic noise which, with suitable amplification, will give rise to appreciable fluctuations in the output voltage. There are no fundamental difficulties in constructing an electronic circuit whose output voltage $U(t)$ will be a random quantity. The values of this quantity may be chosen at widely spaced times $t_1, t_2, \ldots, t_k, \ldots$, so that $U(t_1), U(t_2), \ldots, U(t_k)$ can be regarded as independent. The quantity x_l can now be defined by

$$x_l = \begin{cases} 0 & \text{for} \quad U(t_l) \leqslant a, \\ 1 & \text{for} \quad U(t_l) > a. \end{cases} \tag{1.19}$$

The quantity a is usually referred to as the cut-off level. It should be chosen so that the probability of $x_l = 1$ is 0.5. In fact, the main difficulty lies in the correct choice of a. Various methods are available for this purpose. The simplest method is to take a pair of values x_l and x_{l+1} and construct a quantity y_l in accordance with the rule

$$y_l = \begin{cases} 1, & \text{if} \quad x_l = 1,\ x_{l+1} = 0, \\ 0, & \text{if} \quad x_l = 0,\ x_{l+1} = 1. \end{cases}$$

If, on the other hand, $x_l = x_{l+1}$, then y_l is determined by taking the first of the subsequent pairs x_{l+k}, x_{l+k+1} whose members have different values. If the probability of $x_l = 1$ is denoted by w then the probability that $y_l = 1$ is

$$\frac{w(1-w)}{w(1-w)+(1-w)w} = \frac{1}{2}.$$

It is easy to see that the search for a suitable pair, x_{l+k} x_{l+k+1} will be completed in a finite number of steps, and will occupy, on the average,

$$\frac{1}{2w(1-w)}$$

steps. When $w \approx 0.5$, this will amount to 2 steps.

The second method is particularly convenient when w is close to 0.5. Let $w = 0.5 + \varepsilon$ and let us define y_l by

$$y_l = \begin{cases} 1, & \text{if} \quad x_l \neq x_{l-1}, \\ 0, & \text{if} \quad x_l = x_{l-1}. \end{cases}$$

In this case, the probability that $y_l = 1$ is

$$2w(1-w) = 2\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right) = 2\left(\frac{1}{4}-\varepsilon^2\right) = \frac{1}{2} - 2\varepsilon^2,$$

i.e., it is much closer to 0.5 than is w, if $\varepsilon \ll 0.5$. This method may be iterated by taking 2^m consecutive values of x_l. When a trigger circuit is used as the source of noise, the above method will require the alternate selection of the output signal from different outputs of the trigger, depending on the value of the preceding output signal.

If a particular computer does not incorporate a random-number generator, then use is made of the so-called quasi-random number sequences. A quasi-random sequence $\xi_1, \xi_2, \dots, \xi_n, \dots$ is one defined by a recurrence relation

$$\xi_n = f(\xi_{n-1}, \xi_{n-2}, \dots, \xi_{n-k}) \tag{1.20}$$

which is such that when $1 \leqslant n \leqslant N$ the sequence it generates has the statistical properties of a sequence of independently chosen values of a uniformly distributed random quantity (the distribution need not be uniform, although it must be known).

An example of the generation of quasi-random sequences of the first order ($k = 1$) which was analyzed with a computer is given below. The rule consists of the following. Let ξ_{n-1} be an mth order binary number of the form

$$\xi_{n-1} = \varepsilon_1 2^{-1} + \varepsilon_2 2^{-2} + \dots + \varepsilon_m 2^{-m}.$$

The square of this number is then of the form

$$\xi_{n-1}^2 = \delta_1 2^{-1} + \delta_2 2^{-2} + \dots + \delta_{2m} 2^{-2m}.$$

Let us select the central portion of this number (assuming m is even) and replace (1.20) by

$$\xi_n = f(\xi_{n-1}) = \delta_{\frac{m}{2}+1} 2^{-1} + \delta_{\frac{m}{2}+2} 2^{-2} + \dots + \delta_{\frac{3m}{2}} 2^{-m}.$$

This method may be used to obtain satisfactory sequences for N of the order of a few thousand. Satisfactory quasi-random sequences of considerable length may be obtained by using higher order sequences ($k = 2$ or 3). There are also other methods of constructing quasi-random sequences which are used with electronic computers. Some of these methods are based on operations peculiar to the particular computer (cf. Chapter VI).

4. SOLUTION OF SYSTEMS OF LINEAR ALGEBRAIC EQUATIONS

The solution of systems of linear algebraic equations is of great importance in many fields in which computational mathematics is employed. This is one of the reasons for considering a further class of methods for the solution of linear systems, which are based on the simulation of random processes. One of such methods is described below. Other methods for the solution of systems of linear algebraic equations are described elsewhere [11].

Consider a system of linear equations which may be written in the vector form

$$Ax = b, \tag{1.21}$$

where A is an n-dimensional matrix and x and b are n-dimensional vectors. Suppose that we are dealing with a case where the method of simple iteration is possible, i.e., the matrix A can be written in the form $A = E - B$, where E is the unit matrix and B has characteristic numbers whose moduli are less than unity. The system given by (1.21) is then equivalent to the following "iterative" form:

$$x = Bx + b. \tag{1.22}$$

The solution of (1.21) may be written in the form

$$x = A^{-1}b. \tag{1.23}$$

Subject to the above assumptions, the reciprocal matrix may be expanded into the series

$$A^{-1} = E + B + B^2 + \dots + B^n + \dots \tag{1.24}$$

This series will converge when, and only when, the characteristic numbers of the matrix B are such that

$$|\lambda_i(B)| < 1.^*$$

Let us now substitute (1.24) into (1.23) and write the solution in the form

$$x = b + Bb + B^2b + \dots + B^nb + \dots \tag{1.25}$$

The partial sums of the latter series may be obtained by means of the usual iteration method by successively writing

---

*This is also the condition for convergence of the method of simple iteration.

$$
\begin{aligned}
&x^{(1)} = b + b \\
&x^{(2)} = Bx^{(1)} + b = Bb + b, \quad \ldots, \\
&\ldots\ldots\ldots\ldots \\
&x^{k} = Bx^{(k-1)} + b = B^{k-1}b + B^{k-2}b + \ldots + b, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

The convergence of the method of simple iteration is equivalent to the convergence of the sequence $x^{(1)}, x^{(2)}, \ldots, x^{(k)}, \ldots$

In order to evaluate x let us rewrite these expressions in terms of their coordinates. Let the elements of the matrix B be denoted by B_{ij} so that the mth coordinate x_m of the vector x is

$$
\begin{aligned}
x_m = b_m + \sum_i B_{mi} b_i + \sum_{i_1, i_2} B_{mi_1} B_{i_1 i_2} b_{i_2} + \ldots \\
\ldots + \sum_{i_1, i_2, \ldots, i_r} B_{mi_1} B_{i_1 i_2} \ldots B_{i_{r-1} i_r} b_{i_r}.
\end{aligned}
\tag{1.26}
$$

This expression will be used later. The problem consists in finding a method for evaluating this sum.

To begin with, let us consider the case where the elements of the matrix are positive and the sum of the elements in each column is equal to unity,* i.e.,

$$\sum_i B_{mi} = 1.$$

The quantities B_{mi} can then be regarded as the probabilities for an exhaustive set of incompatible events.

In the preceding section, we were concerned with methods of simulating random events. In the present case, it is necessary to simulate a system of random events.

Consider n independent values $\xi_1, \xi_2, \ldots \xi_n$ which are uniformly distributed in the interval $(0, 1)$. Let us now subdivide this interval into n lengths $B_{m1}, B_{m2}, \ldots, B_{mn}$ $(m = 1, 2, \ldots n)$. If ξ_m lies in the ith section (length B_{mi}) then we shall put $y_m = b_i$. This defines n random quantities $y_1, y_2, \ldots, y_n$. The mathematical expectation of y_m is

$$\mathrm{M} y_m = \sum_{i=1}^{n} B_{mi} b_i,$$

i.e., it is equal to the second term in the series (1.26).

We shall attempt to obtain a random number whose mathematical expectation equals the third term of the series (1.26).

To do this, consider again the quantities ξ_m. If ξ_m lies in the i_1th section, then we will try another value. If ξ_i lies in the i_2 th section, then we shall put $Z_m = b_{l_2}$. This defines the random quantities Z_m. The probability with which the random quantity Z_m assumes the values b_{l_2} is equal to

---

*We are considering this case only to elucidate the probability model. In reality, the series given by (1.25) will diverge under these conditions, since the resulting matrix necessarily has a characteristic value $\lambda = 1$.

$$\sum_{l_1} B_{ml_1} B_{l_1 l_2},$$

and the mathematical expectation of Z_m is given by

$$\mathsf{M} Z_m = \sum_{l_1, l_2} B_{ml_1} B_{l_1 l_2} b_{l_2}.$$

The above construction makes possible a simple generalization whereby it is possible to obtain a random quantity η_m whose mathematical expectation is equal to the sum of the series (1.26), i.e.,

$$\mathsf{M} \eta_m = x_m.$$

Thus, by simulating N times the process of determination of η_m and as a result obtaining the sequence $\eta_m^1, \eta_m^2, \ldots, \eta_m^N$, one can evaluate the arithmetic mean

$$\bar{\eta}_m = \frac{\eta_m^1 + \eta_m^2 + \ldots + \eta_m^N}{N}.$$

which may be taken as an approximation of the required quantity x_m. The error may be estimated by the method described in Section 2.

The random quantity η_m may be constructed as follows. Each element B_{mj} of the matrix B may be written in the form of a product of two factors

$$B_{mj} = F_{mj} P_{mj}, \tag{1.27}$$

where $0 < P_{mj} < 1$.* The components of the vectors on the right-hand side of (1.21) will be written in the form $b_j = f_j p_j$ where $0 < p_j < 1$.**

It will be assumed that

$$p_m + \sum_{j=1}^{n} P_{mj} = 1. \tag{1.28}$$

---

*It will be seen later than when $B_{mj} = 0$ it is convenient to let $P_{mj} = 0$.

**When $b_j = 0$ it is convenient to let $p_j = 0$.

Equation (I.26) can then be written in the form

$$\begin{aligned} x_m = f_m p_m + \sum_i F_{ml} f_l P_{ml} p_l + \cdots \\ \cdots + \sum_{l_1, l_2, \ldots l_r} F_{ml_1} F_{l_1 l_2} \cdots F_{l_{r-1} l_r} f_{l_r} P_{ml_1} \cdots \\ \cdots P_{l_{r-1} l_r} p_{i_r} + \cdots \end{aligned} \tag{1.29}$$

Consider now n subdivisions of the interval [0, 1], each into $n+1$ segments. The lengths of the segments obtained as a result of the mth subdivision will be

$$P_{m1}, P_{m2}, \ldots, P_{mn}, p_m.$$

Let $\xi_0, \xi_1, \ldots, \xi_l, \ldots$ be values of independent, uniformly distributed quantities. The quantity η_m will be defined as follows.

To begin with, consider the mth subdivision of the interval [0, 1] and determine which segment corresponds to ξ_0. If ξ_0 corresponds to the $(n+1)$th segment (length p_m) then we assume that

$$\eta_m = f_m. \tag{1.30}$$

If ξ_0 corresponds to the l_1th part (length P_{ml_1}) then the l_1th subdivision is selected and the segment of this subdivision which corresponds to ξ_1 is noted. Again, if ξ_1 corresponds to the $(n+1)$th segment of the subdivision, then we assume that

$$\eta_m = F_{ml_1} f_{i_1}. \tag{1.31}$$

If, on the other hand, ξ_1 corresponds to the l_2th segment of the subdivision, then ξ_2 is selected and the segment of the l_2th subdivision to which it belongs is noted. This process is continued either until one of the quantities ξ_l falls on the $(n+1)$th segment of the lth subdivision, or indefinitely.

It may, however, be shown that, with a probability equal to unity, the second case (infinite continuation) is never encountered. This follows from the general properties of Markov chains which are discussed in Section 6 of this chapter.

The quantity η_m will thus assume values which are determined by the history of the process. In particular, it assumes the value

$$F_{ml_1} F_{l_1 l_2} \cdots F_{l_{k-1} l_k} f_{l_k} \tag{1.32}$$

with probability

$$P_{ml_1} P_{l_1 l_2} \cdots P_{l_{k-1} l_k} p_{l_k}. \tag{1.33}$$

Thus η_m assumes the values given by (1.32) if ξ_0 lies in the ith segment of the mth subdivision, ξ_1 in the i_2th segment of the l_1th subdivision, and so on, while ξ_k lies in the $(n+1)$th segment of the kth subdivision.

The mathematical expectation of η_m is

$$\mathbf{M}\eta_m = f_m p_m + \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k} F_{mi_1} F_{i_1 i_2} \ldots F_{i_{k-1} i_k} f_{i_k} \cdot \cdot p_{mi_1} p_{i_1 i_2} \ldots p_{i_{k-1}} p_{i_k}. \tag{1.34}$$

This follows from (1.32) and (1.33).

Comparing (1.34) with the expression for the mth component of the unknown vector x, which is given by (1.29), we have

$$\mathrm{M}\eta_m = x_m. \tag{1.35}$$

This identity may be used to calculate x_m by a Monte Carlo simulation of the random quantity η_m.

We have thus arrived at a Monte Carlo process for the solution of systems of linear algebraic equations. There are also other processes of this kind where a statistical model is used for the determination of the reciprocal matrix, or for the solution of (1.21) with an arbitrary matrix A.

The particular advantage of the above method of solution lies in the following. When systems of linear equations are solved by the usual method, the determination of any one of the unknowns requires a knowledge of all the remaining unknowns. In the above method, on the other hand, this is not necessary; each time, a single coordinate x_m is determined. This is shown by the fact that the number of arithmetical operations is proportional to the number of equations and not to the cube of this number as in the standard numerical methods. This is one of the fundamental features of the Monte Carlo method, namely, its suitability for and convenience in the case of multidimensional problems.

5. RANDOM WALK AND THE SOLUTION OF BOUNDARY VALUE PROBLEMS

Boundary value problems and problems with initial conditions for linear differential equations are among the more interesting fields where the Monte Carlo method is applicable. The connection between the solutions of such problems for certain classes of equations, and stochastic processes of the random walk type, has been known for some considerable time [232]. However, the application of this fact to the practical solution of equations became possible only after the development of computers.

We shall see later that the Monte Carlo method is particularly effective in the case of multidimensional problems, since the time necessary for the determination of the value of the solution at a given point depends only on the diameter of the region.

In order to illustrate the basic idea of the method we shall consider the Dirichlet problem for the Laplace equation. Consider a function $f(Q)$ which is defined on the boundary of a simply connected plane region G. It is required to find a function $u(P)$ which satisfies the Laplace equation $\Delta u = 0$ within the given region G and assumes given values $u|_\Gamma = f(Q)$ on the boundary Γ of the region. This problem is usually reduced to a finite difference scheme.

Consider a planar square lattice with a spacing equal to h. We consider only those nodes of the lattice which lie inside the region G. They can be divided into two groups, namely, those which have four neighbors inside G and which are known as internal nodes, and those which have fewer than four neighbors inside G and which are referred to as boundary nodes (Fig. 2).

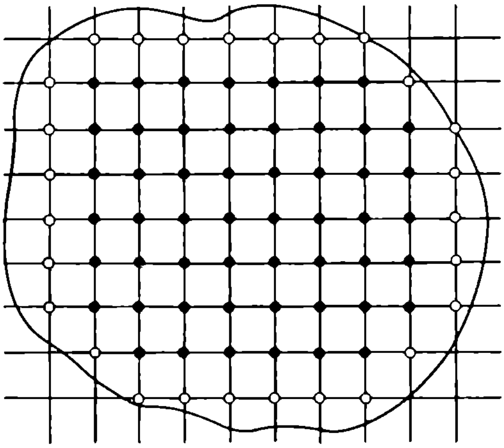


Fig. 2.

The function u assumes the given values $u(Q) = f(Q)$ at the boundary nodes; the values of u on the boundary are attained at the boundary sites by means of special rules.

The values of the function $u(P)$ at the internal nodes will be sought using expressions of the form

$$u(P) = \frac{1}{4}[u(P_1) + u(P_2) + u(P_3) + u(P_4)], \tag{1.36}$$

where P_1, P_2, P_3 and P_4 represent the four sites which are the nearest neighbors of P and lie either inside the region or on the boundary. Equation (1.36) is the usual finite difference system.

Consider the probabilistic scheme related to (1.36). It is usually referred to as the random walk problem. We shall regard the squares of the lattice as blocks in a city and the nodes as crossroads. Suppose that a drunken man leaves a node P and can reach any of the neighboring nodes with equal probability (0.25). Similarly, having reached the next site (crossroads) he can, with equal probability, reach any of the four possible new neighbors. We shall assume that the town is surrounded by a very deep moat, which means that when the man reaches the city boundaries (i.e., a boundary node of the lattice), he will remain at this site by falling into the moat. The problem is to determine the probability that such a man leaving a site P will complete his random walk at the boundary node Q. It may be shown that, with probability equal to unity, he will eventually reach the city boundaries. The required probability cannot easily be obtained in an explicit form. However, it is quite simple to derive a relation involving the probability $u(P, Q)$. We note that the walks which, starting from P, terminate at the point Q are composed of four mutually exclusive classes, namely the walks in which the man reaches Q from P via P_1, P_2, P_3 and P_4 where P_1, P_2, P_3, P_4 denote the four neighbors of P. Since the probabilities of reaching P_i from P are all equal to 0.25, the law of combination of probabilities yields

$$u(P, Q)=\frac{1}{4}\sum_{i=1}^{4} u(P_i, Q). \qquad (1.37)$$

We have thus arrived at the finite difference equation (1.36). Moreover, the probability $u(P, Q)$ satisfies the boundary conditions

$$\left.\begin{aligned} &u(Q, Q)=1, \\ &u(Q', Q)=0 \qquad (Q' \neq Q), \end{aligned}\right\} \qquad (1.38)$$

where Q and Q' are boundary nodes.

It is known that there exists a unique function which will satisfy (1.36) for the given boundary conditions.

By simulating N times the random walk of the man, one can determine the number of times, L, which will result in his reaching Q from P, and this will give an approximate estimate for $u(P, Q)$, namely

$$\frac{L}{N} \approx u(P, Q).$$

We thus have an approximate solution of (1.37) subject to the boundary conditions (1.38).

The probability scheme must be generalized somewhat when it is required to solve the Dirichlet problem with general boundary conditions.

Let us assume that when the man falls into the moat at a boundary node Q, he must pay a fine* $f(Q)$. It is clear that the magnitude of the fine $\xi(P)$ paid by the man starting from the point P is a random quantity.

It may assume the values

$$f(Q_1),\ f(Q_2),\ \dots,\ f(Q_s),$$

where $\{Q_1, Q_2, \dots, Q_s\}$ is the set of all boundary sites. The probability of paying a fine $f(Q_l)$ is equal to $u(P, Q_l)$, so that the mathematical expectation of the fine is

$$w(P) = \mathsf{M}\xi(P) = \sum_{l=1}^{s} f(Q_l)\,u(P, Q_l). \tag{1.39}$$

It is clear that the magnitude of $w(P)$ depends on the starting point P. The function $w(P)$ satisfies the difference equation

$$w(P) = \frac{1}{4}\sum_{l=1}^{4} w(P_l). \tag{1.40}$$

In fact, substituting $Q = Q_l$ into (1.37) and multiplying both sides by $f(Q_i)$ we obtain equation (1.40) after summation over all the boundary sites Q_l.

The function $w(P)$ satisfies the required boundary conditions at the boundary nodes. In fact, substituting $P = Q$ in (1.39) we find that, in view of (1.38), all the components on the right-hand side of (1.39) will vanish except for

$$w(Q) = u(Q, Q)\,f(Q) = f(Q).$$

Thus, the function $w(P)$ assumes the prescribed values on the boundary, i.e., it is a solution of the Dirichlet problem.

Consider now the time necessary for the solution of the problem. Suppose that the coordinates of the site P are (x_0, y_0). The coordinates of the four neighboring sites are then

$$\begin{aligned} P_1 &= (x_0 + 1,\ y_0), \\ P_2 &= (x_0,\ y_0 + 1), \\ P_3 &= (x_0 - 1,\ y_0), \\ P_4 &= (x_0,\ y_0 - 1). \end{aligned}$$

---

*The fine depends on the part of the boundary in which he is found.

Let the current coordinates of the man's position be (x, y), so that initially $x = x_0$, $y = y_0$. When he reaches a neighboring site from P, unity is added to or subtracted from one of the coordinates (x or y). This is continued until he reaches the city boundaries. The process involves the recording of the coordinates (x_0, y_0) and (x, y).

We shall now calculate the number of operations required for the completion of the random walk at the boundary. Suppose that the number of nodes which the man passes through in the ith random walk is ν_i. The time for the solution, T, is then given by

$$T = t(\nu_1 + \nu_2 + \ldots + \nu_N), \tag{1.41}$$

where t is the time spent between successive nodes and N is the total number of simulated random walks which are necessary for a solution of the required accuracy. The quantity N can be determined from the law of large numbers in the usual way, and is given by

$$N = \frac{9\mathrm{D}\xi(P)}{\varepsilon^2} \leqslant \frac{9 \max |f(Q)|^2}{\varepsilon^2}, \tag{1.42}$$

where ε is the error of the solution, $\mathrm{D}\xi(p)$ is the variance of the quantity $\xi(p)$ and $\max|f(Q)|$ is taken over the boundary nodes. This inequality may be obtained from a simple estimate of the variance of $\xi(p)$.

In fact,

$$\mathrm{D}\xi(p) = \mathrm{M}\xi^2(p) - [\mathrm{M}\xi(p)]^2 \leqslant \mathrm{M}\xi^2(p) \leqslant \max_{Q \in \Gamma} |f(Q)|^2,$$

since all values of $|\xi(p)|$ do not exceed $\max|f(Q)|$.

The number of nodes passed through by the man in a random walk is also a random quantity. It follows that the sum of the quantities ν_i is approximately equal to the mathematical expectation of ν multiplied by N, so that the time necessary for the solution of the problem is $T \approx tN(E\nu)$. The mean number of steps in the random walk, $E\nu$, depends on the form of the lattice and the spacing h.

It turns out that the number of steps depends only on the linear dimensions of the lattice, provided it is assumed (as above) that $h = 1$. If r is the radius of the region G, then $E\nu \sim r^2$.* It is significant that this fact holds for a lattice of any number of dimensions, i.e., for any number of independent variables on which the required function $u(p)$ depends. In general, the time necessary for the solution of the problem is $T \sim \frac{tr^2}{\varepsilon^2}$. The accuracy of the solution will in addition depend on the coarseness of the lattice. For example, if it is desired to obtain the solution to an accuracy $\varepsilon = 0.01$ of the maximum value of $f(Q)$, then $r \sim 100$, $N \sim 10^4$ and $T \sim 10^8$ t.

---

*This is shown by Buslenko and Shreider [11].

If it is assumed that the calculations are carried out with an electronic computer for which the time corresponding to a single step is of the order of 100 μsec, then the total time for the solution turns out to be about three hours.

Another important factor is that the solution will involve the storage of a small number of intermediate results, i.e., only the quantities x, y, x_0, y_0 will have to be stored. Specialized electronic computers with small storage capacity and simple construction, which are designed for the solution of boundary value problems by the Monte Carlo method, are based on this principle [11].

The time required to evaluate $w(P)$ at one site is practically independent of the magnitude of the grid. If it is necessary to evaluate $w(P)$ at all the sites, the time of solution of course increases. However, it is often not necessary to evaluate $w(P)$ at all the sites, but only at certain critical ones.

This problem also illustrates the basic feature of the Monte Carlo method, namely, its suitability for multidimensional problems.

The generalized random walk problem is related to the solution of the general linear second-order elliptic equation

$$a(x, y)\frac{\partial^2 u}{\partial x^2} + 2b(x, y)\frac{\partial^2 u}{\partial x\,\partial y} + c(x, y)\frac{\partial^2 u}{\partial y^2} + d(x, y)\frac{\partial u}{\partial x} + e(x, y)\frac{\partial u}{\partial y} + f(x, y)u = 0.$$

Here, the probability that the man will reach a given site from another depends on the site at which he is found at a particular time.

It is also possible to consider more general boundary conditions. Thus, a condition of the form

$$\frac{\partial u}{\partial n}\Big|_\Gamma = ku\Big|_\Gamma + f(Q)$$

will be obtained if the man can, with a certain probability, leave the moat surrounding the city and then continue the random walk.*

---

*The random walk process has a special form when $\frac{\partial u}{\partial n}\Big|_\Gamma = 0$. This is the so-called Neumann problem, and is discussed in Section 6.

As an example of a nonstationary problem, consider the equation of thermal conductivity

$$\frac{\partial u}{\partial t}=\Delta u, \tag{1.43}$$

whose solution is a function u which depends on the spatial coordinates and on time. Let us assume that a lattice having a spacing h is inscribed in the region D in which it is required to find the solution $u(P, t)$. This function should satisfy the boundary condition

$$u|_{\Gamma}=f(Q) \tag{1.44}$$

and the initial condition

$$u|_{t=0}=g(P). \tag{1.45}$$

Consider the sequence

$$t=0,\ 1,\ 2,\ \ldots,\ k,\ \ldots$$

By choosing a suitable relation between the time scale and the lattice constant one can obtain the equation

$$u_k(P)=\frac{1}{4}\left[u_{k-1}(P_1)+u_{k-1}(P_2)+u_{k-1}(P_3)+u_{k-1}(P_n)\right]. \tag{1.46}$$

where P_1, P_2, P_3, P_4 are the four nodes neighboring P. The Laplace equation is obtained by removing the subscript k from (1.46).

Consider now the following random process whose purpose is to find the value of $u_k(P)$ at the point P at time k.

The lattice used in the preceding problem can again be employed, except that the man will now take exactly one unit of time to cover the distance between neighboring sites.

We shall assume that he starts from a crossroads P, reaches one of the neighboring crossroads with probability 0.25 and continues in this way until he reaches the boundary and remains there. The entire process is allowed to continue for not more than k steps. If after k steps the man does not fall into the moat, but reaches an internal site P, then he is required to pay a fine $\xi=g(P)$. If, on the other hand, he does fall into the moat within the time, then he is required to pay a fine $\xi=f(Q)$, where Q is the boundary point at which he falls into the moat.

Altogether there will be N such random walks, and the total fine divided by N is the approximate solution of the finite difference problem for the thermal conductivity equation (1.46) satisfying the conditions (1.44) and (1.45). In order to show this we shall calculate the mathematical expectation of the fine.

Let $v_k(P, Q)$ be the probability that a man leaving P will reach a general point Q after k steps. It is easy to show that this probability will satisfy the following condition:

$$v_k(P, Q) = \frac{1}{4}[v_{k-1}(P_1, Q) + v_{k-1}(P_2, Q) + v_{k-1}(P_3, Q) + v_{k-1}(P_4, Q)]. \tag{1.47}$$

The quantity $v_k(P, Q)$ satisfies the boundary conditions

$$\left.\begin{aligned} v_k(Q, Q) &= 1, \\ v_k(Q, Q') &= 0 \end{aligned}\right\} \tag{1.48}$$

(when $Q \neq Q'$, where Q' and Q are boundary nodes), and the initial conditions

$$v_0(P, P) = 1, \quad v_0(P, Q) = 0,$$

where P and Q are two distinct nodes (the node P is an internal node, while Q is a general node).

Consider the mathematical expectation of the fine paid by the man starting from the point P. The magnitude of the fine assumes the values $g(P_1), g(P_r), \ldots, g(P_2), f(Q_1), f(Q_2), \ldots, f(Q_s)$. It follows that the mathematical expectation of the fine is

$$w_k(P) = E\xi(P) = \sum_{l=1}^{r} v_k(P, P_l)\, g(P_l) + \sum_{j=1}^{s} v_k(P, Q_j)\, f(Q_j). \tag{1.49}$$

Let us substitute $Q = Q_j$ into (1.47) and then multiply both sides by $f(Q_j)$. Next, substitute $Q = P_l$ into (1.47) and multiply by $g(P_l)$. Addition of all the products yields

$$w_k(P) = \frac{1}{4}\sum_{l=1}^{4} w_{k-1}(P_l), \tag{1.50}$$

i.e., the mathematical expectation of the fine is the solution of the finite difference equation of thermal conductivity (1.46). Boundary and initial conditions for $w_k(P)$ can easily be verified with the aid of (1.48). In fact, let $P = Q_j$, where Q_j is a boundary node. The right-hand side of (1.49) will then include the single nonzero term

$$w_k(Q_j) = v_k(Q_j Q_j)\, f(Q_j) = f(Q_j). \tag{1.51}$$

If it is assumed that $P = P_l$ and $k = 0$, then in (1.49) there will also be a single term, namely,

$$w_0(P_i) = v_0(P_i, P_i)\,g(P_i) = g(P_i). \tag{1.52}$$

Since the solution of (1.50) which satisfies (1.51) and (1.52) is unique, it follows that the above method may be used to obtain it.

A further important feature of the above method is the fact that it may be used to solve the problem by determining the values of w at any one given point, whereas in the solution of equation (1.36) by the usual iteration method the values of w must be found at all the lattice sites. It turns out that the time necessary for the determination of a single value of $w(P)$ is independent of the number of independent variables in the problem.

6. THE MONTE CARLO METHOD AND THE SIMULATION OF MARKOV PROCESSES IN A COMPUTER

The general mathematical setup of the Monte Carlo method can be described in terms of the so-called Markov processes. Only discrete Markov processes with a finite set of states (see [52]), called Markov chains, are considered below.

A system S, possessing a finite set of states $M\{s_1, s_2, \ldots, s_l\}$, is called a Markov chain. At any discrete time instant $t = 0, 1, 2, \ldots, n$ the system S is in one definite state s_i.

The state s_i is associated with a set of conditional probabilities $p_{i1}, p_{i2}, \ldots, p_{il}$. The quantity p_{ij} is the probability that the system, being at the nth time instant in the state s_i, passes into the state s_j at the $(n+1)$th instant. In other words, p_{ij} is the probability of the transition $s_i \to s_j$. An essential feature of the situation is that the probability of transition depends only on the initial state s_i, and is independent of the past history of the system. This constitutes the "Markovian" property of the process.

The totality of all conditional probabilities p_{ij} forms a matrix $P = (p_{ij})$ which completely defines the properties of the given system. The state s_i is called absorbing if the probability that the system S remains in this state, having once passed into it, is one. In terms of conditional probabilities this means that

$$p_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

The state s_i is said to be connected with the state s_j if the probability that the system S, having passed into the state s_i, passes into the state s_j in some finite number of time intervals k is different from zero.

A Markov chain is said to be terminating if all of its states are connected with some absorbing state. The following theorem may serve as a justification for such a term.

Theorem. If a Markov chain is terminating, then with probability unity the system S passes into one of the absorbing states in a finite number of time intervals.

Proof. We shall denote by $q_i(t)$ the probability that in time t the system passes from the state s_i into one of the absorbing states. It is clear that $q_i(t)$ can only increase with an increase of t.

Then in accordance with the condition there exists a value $t=t_0$ such that for all i we have $q_i(t_0)>0$. That is to say, every state of the system, with nonzero probability, can pass into an absorbing state in time t_0.

We shall write

$$q=\min_i q_i(t_0). \tag{1.53}$$

It is clear that $q>0$. The probability, that in time t_0 the system does not pass into an absorbing state, does not exceed the quantity $1-q$.

By virtue of the fact that the probability of transition does not depend on the past history, the probability of the system S not passing into an absorbing state in time νt_0 (an integral multiple of t_0) does not exceed

$$(1-q)^\nu.$$

When $\nu\to\infty$ this quantity tends to zero.

This means that with probability unity the system S passes into an absorbing state in a finite time.

Note. We shall denote by τ the lifetime of the system, that is to say, the time it takes to reach an absorbing state. The time τ is a random quantity, the mathematical expectation of which is evaluated in terms of the quantity q.

Indeed, if the initial state is s_i, then the mathematical expectation

$$\begin{aligned}\mathrm{M}\tau &= \sum_{t=1}^{\infty} t\,(q_i(t)-q_i(t-1))\\ &\leqslant \sum_{t=1}^{\infty} t\,(1-q_i(t-1)) \leqslant \sum_{\nu=1}^{\infty}\sum_{t=\nu t_0+1}^{(\nu+1)t_0} t\,(1-q_i((\nu+1)t))\\ &\leqslant t_0\sum_{\nu=1}^{\infty}(\nu+1)(1-q)^\nu=\frac{t_0}{q^2}.\end{aligned}$$

The sum of the last series is calculated by differentiating term-by-term the geometrical progression, whence the result

$$\mathrm{M}\tau\leqslant\frac{t_0}{q^2} \tag{1.54}$$

is obtained. It is clear from (1.54) that the mean time of transition into an absorbing state is greater when the probability of transition into the absorbing state is smaller.

All algorithms for solving problems by the Monte Carlo method which have been considered above and will be discussed in the remaining chapters of this book can be described in the following manner.

Some terminating Markov chain is being simulated. The process of successive transitions of this chain from state to state takes place in accordance with the scheme

$$X \sim s_{i_0} \rightarrow s_{i_1} \rightarrow s_{i_2} \rightarrow \ldots \rightarrow s_{i_t}, \qquad (1.55)$$

where s_{i_t} is one of the absorbing states. A certain function $F(X)$ of the sequence of transitions (1.55) is determined thereby. The function $F(X)$ is a random quantity. After the value of $F(X)$ has been found, the system S returns to the initial state s_{i_0} and the process of transitions begins anew. Altogether N independent runs of the given Markov chain from the state s_{i_0} into one of the absorbing states are carried out. As a result the sum

$$\frac{1}{N}\sum_{X} F(X), \qquad (1.56)$$

taken over all the sequences of transitions (1.55) which have been carried out, is obtained. The sum (1.56) approaches the quantity $\mathbf{M}\,F(X)$ whose value is required in the given problem. The total time required for solution of the problem is

$$T \approx N(\mathbf{M}\,\tau)\,T_0, \qquad (1.57)$$

where $\mathbf{M}\tau$ is the mathematical expectation of the number of transitions in the sequence (1.55), while T_0 is the mean time for carrying out one of the transitions in the computer.

Thus in the setup for the solution of a system of linear equations, which was considered in Section 4, the fundamental Markov chain has $n+1$ states. The states correspond to divisions of a closed interval. The probability of the transition from the ith state to the jth is P_{ij} (with $i \leqslant n,\ j \leqslant n$). The probability of the transition from the ith state to an absorbing one is p_i. The initial state is denoted by the number m. The function $F(X)$ is defined as

$$F(X) = F_{mi_1} F_{i_1 i_2} \ldots F_{i_{k-1} i_k} f_{i_k}.$$

In the setup for the random walk problem (Section 5), the states of a Markov chain are identified with the nodes of a lattice. The passage of a wandering particle from node to node corresponds to the transition of the system from state to state. The boundary nodes correspond to the absorbing states of the chain, since no more transitions take place after the point has arrived at the boundary. The initial state corresponds to the node from which the random walk is commenced. The function $F(X)$ in this case depends only on the number of the absorbing state (a boundary site), and not on the history of the random walk. One could use the relationship (1.54) for the evaluation of the wandering time, but in the given case a better method was derived in Section 5. It is clear that the lifetime of the system depends on the dimensions of the region. If the initial point lies far from the boundary, then only after many steps can the walk terminate at the boundary.

In the case where a problem with initial and boundary conditions (for example, the equation of heat conduction) is being solved, the states of the Markov chain correspond to the nodes of a space-time lattice. In other words, the state of the Markov chain is defined by the pair (P, t), where P is the node of the space lattice, and t is the time elapsed since the commencement of the random walk.

Pairs such that P is a boundary node, and also pairs such that $t = k$* ,are absorbing states. That is, after k steps the process is terminated if the wandering point has not reached the boundary earlier. In this case also $F(X)$ is determined by the number of the absorbing state alone, and not by the history of transitions.

Side by side with terminating Markov chains, in the Monte Carlo method it is also possible to use chains of another kind, the so-called ergodic Markov chains (see [52]). Ergodic Markov chains do not possess absorbing states, and for any pair of states s_i and s_j there exists a number of steps k such that the probability of transition from s_i to s_j in k steps is different from zero. In this case there exists, for each state s_i, a limiting probability p_i of the system passing into this state. This means that for a sufficiently large number of steps N the number of cases M_i where the system is found in the state s_i satisfies the condition

$$\frac{M_i}{N} \approx p_i \tag{1.58}$$

independently of the initial state of the system. The larger N is, the more accurately condition (1.58) is fulfilled. In other words, (1.58) may be considered as a generalized version of the law of large numbers for the case of dependent trials. A more commonly found version of this law is as follows. Let $\Phi(s)$ be a function depending on the state of the Markov chain* and let the history of the process for N steps be described by the sequence of transitions

---

*If the value $v_k(P)$ is being sought.

$$X \sim s_{l_0} \rightarrow s_{l_1} \rightarrow s_{l_2} \rightarrow \ldots \rightarrow s_{l_N}.$$

Then with probability arbitrarily close to unity, for a sufficiently large N, we have the approximate equality

$$\frac{1}{N}\sum_{k=1}^{N} \Phi(s_{l_k}) \approx \sum \Phi(s_l) p_l. \tag{1.59}$$

The relationships (1.58) and (1.59) make possible the use of ergodic Markov chains for solving problems by the Monte Carlo method. (Obviously, in practice it is possible to take only a finite sequence of transitions, fixing the quantity N in advance.)

An example of the use of ergodic Markov chains is obtained if we consider the problem of the random walk in which, when the point reaches the boundary, the process does not terminate, but the point is "reflected" from the boundary.**

In contrast to that considered in Section 5, this version of the random walk process is connected not with the problem of Dirichlet but with that of Neumann (the second boundary value problem). This problem consists of finding the solution of the equation $\Delta u = 0$, which satisfies the condition $\left.\frac{\partial u}{\partial n}\right|_\Gamma = 0$ (the derivative with respect to a normal vector at the boundary is zero). In order to determine approximately the value of the required function $u(P)$ at some interior node P, it is necessary to carry out the random walk process which has been described, and determine how many times the wandering point passes through the node P. If the number of "hits" is MP, and the total number of steps (transitions) is N, then $\frac{MP}{N} \approx u(P)$.

It will be observed that it is possible at the same time to note the number of passages through several nodes, and thus determine MP at once for several P. We note that the second boundary value problem defines the required function $u(P)$ with an accuracy up to a constant. The method considered determines the function $u(P)$ which satisfies the additional condition

---

*It is possible to consider also certain classes of functions which depend on transition sequences.

**That is to say, in the interpretation suggested in Section 5, the city is surrounded by a wall instead of a moat; on stumbling against it the man is turned back

$$\sum_{P} u(P) = 1, \tag{1.60}$$

where the sum is taken over all the interior nodes of the lattice. The condition (1.60) follows from the obvious condition

$$\sum p_i = 1,$$

where p_i are the limiting probabilities for transitions into the states of the Markov chain (i.e., for the wandering point to pass through the interior nodes of the lattice).

The example presented above indicates that ergodic Markov chains also can be used for statistical simulation, just as can terminating chains, although at present the latter are used in the overwhelming majority of cases.

We note that by using a digital computer we cannot in principle obtain anything beyond the scope of finite Markov processes. Therefore, if there arises a need to simulate a continuous stationary process, for example, then this will be successful only insofar as the process can well be approximated by a Markov chain.

Indeed, the operation of a computer solving the usual type of problem can be depicted as a process of successive transitions of a finite automatic machine from state to state.

At each step the state of the machine is determined by the contents of the registers of the machine and of the locations of the memory. For simplicity, we shall suppose that after the entry of the initial data into the machine no new information is given to the machine. Then the state of the machine at the following step is fully determined by its previous state. Now let the machine receive information at each step from the random-number generator. Then the next state of the machine is determined not only by the preceding value, but also by a certain random quantity. Thus the probability of transition into one new state or another is determined by the preceding state and by the probability distribution of a random quantity. Since usually the values given by the random-number generator at each step are independent, then the transition probabilities are determined only by the preceding states, and not by the past history of the process.

If we could simulate in the random-number generator complicated processes with an after-effect, then the possibilities of the machine would be extended. At present such an after-effect in random-number generators, in view of its irregularity, is a troublesome phenomenon which has to be avoided.

It was mentioned in Section 1 that the most successful (and most common) applications of the Monte Carlo method are to those problems which by their very nature are connected with stochastic processes. The most important part of the solution in this case consists of the choice of a finite Markov chain which is a reasonable approximation to the process. As a rule, this problem is reduced to the approximate representation of some very complicated Markov chain by a simpler Markov chain. In other words, there is a system S with a set of states $\{s_1, s_2, \ldots, s_k\}$, a transition probability matrix (p_{ij}) and a function $F(X)$ depending on the transition sequence. It is required to form a system S^* with a smaller set of states $s_1^*, s_2^*, \ldots, s_l^*$, a simpler transition probability (p_{ij}^*), a more easily calculated function $F^*(X^*)$ and (as far as possible) a shorter mean lifetime $\mathsf{M}\tau^*$ (for a terminating chain), so that

$$\mathsf{M}F^*(X^*) \approx \mathsf{M}F(X). \tag{1.61}$$

We shall consider the basic methods underlying such an approximation.

The first method consists of uniting some of the states. In a number of cases it is expedient to unite certain states of the system S. In this case the states of the system S^* are classes of states of the system S. This is possible if

a) the function $F(X)$ varies only slightly, if in the sequence X of states individual states are replaced by states belonging to the same class;

b) the probability p_{iK} that the state s_i passes into one of the states of a given class K is nearly the same for all states s_i of one and the same class H. It is then possible to set $p_{iK} \approx p_{HK}$, the probability of the transition from class H into class K.*

Problems concerning particles passing through matter may serve as an example of such a uniting process. Here the state of a particle is specified by the three position coordinates and velocity vectors (x, y, z, v_x, v_y, v_z). However, in a number of problems, as, for example, where particles pass through a flat layer, states with the same values of x and v_x can be taken as equivalent (and thus can be united in a single class).

The second method consists of replacing the probability matrix (p_{ij}) by one similar to it, in which very small values p_{ij} are replaced by zeros.

---

* It is also possible to take

$$p_{HK} = \frac{1}{r} \sum_i p_{iK},$$

where the averaging process is carried out over the r states forming the class H.

The third method involves an alteration of the time scale. In this method the system S is replaced by a system S^* having the same states, but considered only at time instants that are multiples of ν:

$$0, \nu, 2\nu, \ldots$$

In this case the transition matrix (p_{ij}) is replaced by the νth power of the same matrix.

The fourth method consists of approximating the function $F(X)$ by a function $F^*(X^*)$. For example, in the solution of the Dirichlet problem (see Section 5) the function $f(Q)$ at the boundary nodes can be replaced by a polynomial which approximates it in the mean.

There are also certain more complicated methods which in essence are combinations of those just described.

As yet there exists no general theory permitting effective approximation of a complicated process by means of a finite Markov chain as is done for the case of one-dimensional integrals by means of the well-known quadrature formulas. Normally, specific methods of simplification are used for each class of problem. To a large extent these methods form the contents of Chapters II and VI.

Chapter II

Evaluation of Definite Integrals

The evaluation of multiple integrals is one of the more important fields of application of the Monte Carlo method. A large number of quadrature formulas are available for the evaluation of single integrals. However, the number of such formulas which may be used for multidimensional integrals is relatively small. Moreover, the actual calculation involved in such quadratures is rather difficult. The integrals can often be more conveniently evaluated by the less accurate though simpler devices of the Monte Carlo method.

1. SIMPLE APPLICATIONS OF THE MONTE CARLO METHOD

In this section we shall be concerned with two simple applications of the Monte Carlo method which may be used to evaluate integrals of the form

$$J = \int_a^b f(x)\,dx.$$

In the first approach the average value of $f(x)$ is calculated, and in the second use is made of the geometrical interpretation of the integral, which is looked upon as an area.

Methods of estimating the accuracy and efficiency of the Monte Carlo method will also be given.

1. Calculation of the Mean Value of a Function

Let ξ be a random quantity which is uniformly distributed in the interval (a, b) with a probability density $p_\xi(x)$ defined by

$$p_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } x \text{ is in } (a, b), \\ 0 & \text{in the opposite case.} \end{cases}$$

The mathematical expectation of $f(\xi)$ is given by

$$\mathsf{M}f(\xi)=\int_a^b f(x)\,p_\xi(x)\,dx=\frac{J}{b-a}$$

Suppose that as a result of N trials the values of ξ obtained were $\xi_1, \xi_2, \ldots, \xi_N$. Since for large N

$$\mathsf{M}f(\xi)\approx\frac{1}{N}\sum_{i=1}^{N} f(\xi_i),$$

it follows that the quantity

$$\theta_1=\frac{b-a}{N}\sum_{i=1}^{N} f(\xi_i) \tag{2.1}$$

is an approximate estimate of the integral J.

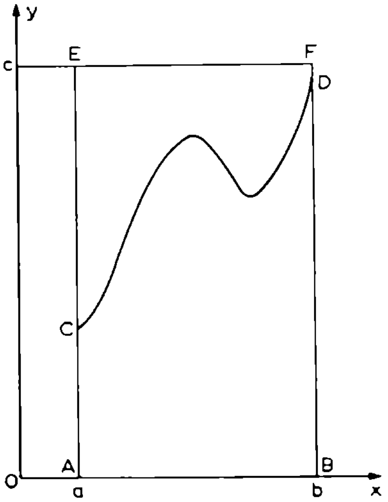


Fig. 3.

In practice, the ξ_i are first used to calculate $f(\xi_i)$, the sum $f(\xi_1)+f(\xi_2)+\ldots+f(\xi_i)$ is then evaluated, and after the completion of all the N trials, θ_1 is determined from the above formula.

2. Area Calculations

For simplicity, consider a bounded integrand $0\leqslant f(x)\leqslant c$ and suppose that (ξ, η) is a random point which is uniformly distributed in the rectangle R

$$R=\{a<x<b,\ 0<y<c\}.$$

The corresponding probability density $p_{\xi\eta}(x, y)$ is given by

$$p_{\xi\eta}(x, y)=\begin{cases}\dfrac{1}{c(b-a)}, & \text{if } (x, y) \text{ belongs to } R,\\ 0 & \text{in the opposite case.}\end{cases}$$

Suppose that the following N points belonging to R have been found: $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_N, \eta_N)$. It is geometrically obvious (Fig. 3) that if N' of these points lie under the curve $y=f(x)$, then the ratio of the areas $ABDC : ABFE \approx N' : N$, or

$$\frac{\int\limits_a^b f(x)\,dx}{c(b-a)} \approx \frac{N'}{N}.$$

The following quantity may therefore be taken as an approximate estimate of the integral:

$$\theta_2=c(b-a)\frac{N'}{N}. \tag{2.2}$$

In practice, the condition $\eta_i < f(\xi_i)$ is checked for each point (ξ_i, η_i): if this condition is satisfied then unity is added to the N' counter; if it is not satisfied, the count remains unchanged. The quantity θ_2 is evaluated after all the N trials have been completed.

It will be shown later that the estimator (2.2) is, as a rule, not as good as (2.1). Nevertheless, many authors (for example, Householder [63] and Kitov and Krinitskii [39]) use only (2.2) for the evaluation of integrals by the Monte Carlo method.

3. Statistical Estimate of the Accuracy

The accuracy of the Monte Carlo method was considered in Section 2 of Chapter I.

In what follows we shall assume that the simulated random quantity θ follows the almost-normal distribution

$$\mathsf{P}\{\theta < x\} \approx \Phi\left(\frac{x-\mathsf{M}\theta}{\sqrt{\mathsf{D}\theta}}\right),$$

where $\Phi(x)$ is given by

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt,$$

and can be used to estimate the probability of deviation of θ from the mathematical expectation $\mathsf{M}\theta$.

Consider an arbitrary degree of reliability p and let x_p be the root of the equation

$$2\Phi(x_p) - 1 = p.$$

The probability of the inequality

$$|\theta - \mathsf{M}\theta| \leqslant x_p \sqrt{\mathsf{D}\theta}$$

is approximately equal to p.

The value $x_p = 3$, which corresponds to $p > 0.99$, is frequently recommended. However, in practice, this is found to lead to an overestimate of the amount by which θ differs from $\mathsf{M}\theta$, and therefore many authors use $p = 0.95$, i.e., a 95% level of reliability, which corresponds to $x_p = 1.96$.

The probable error is defined by

$$\delta_{\text{prob}} = 0.675 \sqrt{\mathsf{D}\theta}.$$

The value $x_p = 0.675$ corresponds to a 50% level ($p = 0.5$) so that the probabilities of the inequalities

$$|\theta - \mathsf{M}\theta| \leqslant \delta_{\text{prob}} \text{ and } |\theta - \mathsf{M}\theta| \geqslant \delta_{\text{prob}}$$

are equal. The probable error δ_{prob} is often used in practice to characterize the order of magnitude of the error. The standard error

$$\sigma = \sqrt{\mathsf{D}\theta}$$

is also used for this purpose.

The majority of the estimates used in evaluating integrals by the Monte Carlo method take the form of the arithmetic mean

$$\theta = \frac{1}{N} \sum_{i=1}^{N} \zeta_i$$

of similar random quantities ζ_i whose mathematical expectation is equal to the required value of the integral:

$$\mathsf{M}\zeta_i = J.$$

Let $\mathsf{D}\zeta_i = \mathsf{D}\zeta$. The mathematical expectation and variance of the random quantity θ are given by

$$\mathsf{M}\theta = J, \quad \mathsf{D}\theta = \frac{\mathsf{D}\zeta}{N}, \text{ respectively.}$$

It follows from the central limit theorem of the theory of probability [23] that for large N the distribution of θ is approximately normal. Hence, with a probability approximately equal to p, we have

$$|\theta - J| \leqslant x_p \sqrt{\frac{\mathrm{D}\zeta}{N}}.$$

In practice it is very important that it should be possible to estimate the variance $\mathrm{D}\zeta$ easily in the course of the calculation of θ. To achieve this, it is sufficient to compute the square of ζ_i in addition to ζ_i^2 itself:

$$\mathrm{D}\zeta \approx \frac{1}{N-1}\left(\sum_{i=1}^{N} \zeta_i^2 - N\theta^2\right).$$

Since a high accuracy is not required for $\mathrm{D}\zeta$ (it is sufficient to have one or one and a half significant figures), it follows that for $N > 30$ it is possible to use the simpler formula

$$\mathrm{D}\zeta \approx \frac{1}{N}\sum_{i=1}^{N} \zeta_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} \zeta_i\right)^2.$$

4. The Accuracy of Simple Methods

Two different methods of evaluating the integral

$$J = \int_a^b f(x)\,dx$$

were given in paragraphs 1 and 2. These methods led to the estimates θ_1 and θ_2. The accuracy of these estimates can be compared by evaluating their variances.

The estimate given by (2.1) is the arithmetic mean of similar quantities $\zeta^{(1)} = (b-a) f(\xi)$. Hence,

$$\mathrm{D}\theta_1 = \frac{\mathrm{D}\zeta^{(1)}}{N},$$

where

$$\mathrm{D}\zeta^{(1)} = (b-a)\int_a^b f^2(x)\,dx - J^2.$$

The estimate given by (2.2) can also be looked upon as the arithmetic mean of similar quantities $\zeta^{(2)} = c(b-a)\,g(\xi, \eta)$ where

$$g(x, y)=\begin{cases}1, & \text{if} \quad y<f(x), \\ 0, & \text{if} \quad y \geqslant f(x).\end{cases}$$

Hence,

$$\mathsf{D}\theta_2=\frac{\mathsf{D}\zeta^{(2)}}{N},$$

where*

$$\mathsf{D}\zeta^{(2)}=c(b-a)J-J^2.$$

Since by definition $0 \leqslant f(x) \leqslant c$, it follows that

$$\int_a^b f^2(x)\,dx \leqslant cJ.$$

Hence, in all cases

$$\mathsf{D}\zeta^{(1)} \leqslant \mathsf{D}\zeta^{(2)}.$$

Example. Evaluate the integral

$$J=\int_0^1 e^x\,dx.$$

The exact value of this integral is known to be

$$J=e-1=1.71828\ \ldots$$

Using the estimate (2.1) we have the following approximation to J

$$J_l=\frac{1}{N}\sum_{l=1}^{N} e^{\gamma_l},$$

where $\gamma_1, \gamma_2, \ldots$ are values of a random quantity which is uniformly distributed in the interval $(0, 1)$. The variance of the averaged quantity is

$$\mathsf{D}\zeta^{(1)}=\frac{1}{2}(e^2-1)-(e-1)^2=0.2420.$$

---

*The same result can easily be obtained when a binomial distribution is assumed for $\zeta^{(2)}$.

The estimate (2.2) can also be used to evaluate the integral by setting $c = e$. This yields*

$$J_2 = \frac{e}{N} \sum_{i=1}^{N} g(\gamma_{2i-1}, \gamma_{2i}),$$

where

$$g(x, y) = \begin{cases} 1 & \text{for } ey < e^x, \\ 0 & \text{for } ey \geqslant e^x. \end{cases}$$

The variance of the averaged quantity is

$$\mathrm{D}\zeta^{(2)} = eJ - J^2 = 1.7183.$$

Numerical details of this calculation are given at the end of Section 2.

5. Estimate of the Efficiency

Consider any two ways of evaluating the integral J by the Monte Carlo method. It is natural to assume that the more efficient method is that which will allow us to reach a given accuracy more rapidly.

Let θ' and θ'' be the estimates obtained by the two methods, so that

$$\mathrm{M}\theta' = \mathrm{M}\theta'' = J;$$

$$\mathrm{D}\theta' = \frac{\mathrm{D}\zeta'}{N}; \quad \mathrm{D}\theta'' = \frac{\mathrm{D}\zeta''}{N}.$$

Let τ' and τ'' be the times necessary to calculate the quantities ζ' and ζ''. Since the two methods are only to be compared, it may be assumed that τ is in some arbitrary units. The times necessary to achieve a given accuracy are proportional to the products $\tau' \cdot \mathrm{D}\zeta'$ and $\tau'' \cdot \mathrm{D}\zeta''$. It follows that the efficiency of the method is inversely proportional to the product $\tau \mathrm{D}\zeta$. The quantity τ (and also $\mathrm{D}\zeta$) may be estimated empirically from a small number of trials. It was shown in paragraph 4 that

$$\mathrm{D}\zeta^{(1)} \leqslant \mathrm{D}\zeta^{(2)}$$

and hence, θ_1 is, as a rule, a better estimate than θ_2.

However, it may turn out that $\tau_2 \ll \tau_1$ (for example, if the function $y = f(x)$ is defined by an algebraic equation $F(x, y) = 0$ for which an explicit solution cannot be found); the method involving (2.2) will then be more convenient than that involving (2.1).

---

*$g(\gamma_{2i-1}, \gamma_{2i})$ can be used instead of $g(\gamma_i, \gamma_{N+i})$. The quantity η is uniformly distributed in the interval $(0, e)$ and is obtained by multiplying γ by e.

2. SOME METHODS FOR REDUCING THE VARIANCE

The rate of convergence of the ordinary Monte Carlo method is relatively low and is of the order of $N^{-\frac{1}{2}}$. In order to increase the accuracy by a factor of 10, the number N, i.e., the volume of the calculations, must be increased by a factor of 100. It is clear that an appreciable increase in the accuracy cannot be obtained in this way. The following methods for reducing the variance, and thus improving the accuracy without increasing the number of trials, are therefore of particular importance.

6. Separation of the Principal Part

Simple Monte Carlo methods are subject to relatively large errors. It is therefore very convenient, whenever possible, to determine the main part of the final result by ordinary methods and use the Monte Carlo method to estimate the correction. This principle holds for all problems solved by the Monte Carlo method.

Consider its application to the evaluation of the integral

$$J = \int_a^b f(x)\,dx$$

and suppose that we have found a function $g(x) \approx f(x)$ whose integral is known, so that

$$\int_a^b g(x)\,dx = I.$$

Since the mathematical expectation of the function

$$\zeta^{(3)} = (b-a)[f(\xi) - g(\xi)] + I$$

is equal to J as before, it follows that it may be approximately estimated from

$$\theta_3 = \frac{b-a}{N} \sum_{i=1}^{N} [f(\xi_i) - g(\xi_i)] + I, \tag{2.3}$$

where $\xi_1, \xi_2, \ldots$ are values of a random quantity ξ which is uniformly distributed in (a, b).

The variance is given by

$$\mathsf{D}\zeta^{(3)} = (b-a)\int_a^b [f(x)-g(x)]^2\,dx - (J-I)^2,$$

and when $|g(x)-f(x)| \to 0$, then $\mathsf{D}\zeta^{(3)} \to 0$.

It is particularly useful to separate out the principal part when the Monte Carlo method is used to evaluate areas (or volumes) with the aid of (2.2).

Suppose that the area $ABDC$ in Fig. 4 is equal to J, and let the areas of the rectangles $ABFE$ and $ABB'A'$ be equal to S and S', respectively.

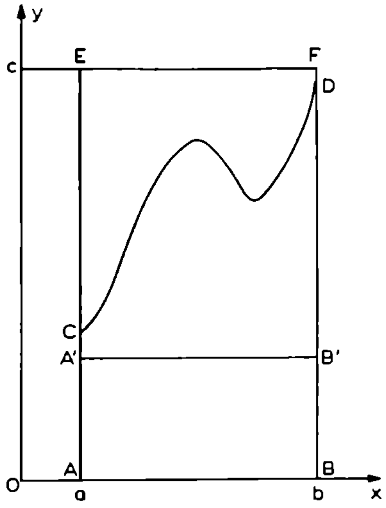


Fig. 4.

By choosing random points in $ABFE$, and counting the number of them which lie in $ABDC$, we obtain for the variance

$$\mathsf{D}\zeta = J(S-J).$$

If, on the other hand, the random points are selected in $A'B'FE$, and the number of them which lie in $A'B'DC$ is determined, then another value is obtained for the variance:

$$\mathsf{D}\zeta' = (J-S')(S-J).$$

Example. Evaluate the integral

$$J = \int_0^1 e^x\,dx.$$

Since $e^x = 1 + x + \ldots$, we shall take $g(x) = 1 + x$ as the principal part. According to (2.3) we have

$$J_3 = \frac{1}{N}\sum_{l=1}^{N}\left(e^{\gamma_l} - \gamma_l\right) + \frac{1}{2},$$

where $\gamma_1, \gamma_2, \ldots$ are the values of a random quantity which is uniformly distributed in (0, 1). The variance of the averaged quantity is then given by

$$\mathsf{D}\zeta^{(3)} = \frac{1}{2}(e-1)(5-e) - \frac{23}{12} = 0.0437.$$

It was shown earlier that when this integral is evaluated with the aid of (2.1) and (2.2), the corresponding variances are 0.2420 and 1.7183, respectively. Numerical details for the above example are given in paragraph 11.

7. Selective Sampling

This method is based on a variable sampling procedure [173-175, 38, 193] in which a larger number of points is selected in the "more important" parts of the region of integration.

Let ξ be a random quantity with a probability density $p(x) > 0$, $a < x < b$, which is normalized so that $\int_a^b p(x)\,dx = 1$. The required integral may be rewritten in the form

$$J = \int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{p(x)}\,p(x)\,dx.$$

Consider the function

$$\zeta^{(4)} = \frac{f(\xi)}{p(\xi)}$$

whose mathematical expectation is

$$\mathsf{M}\zeta^{(4)} = J.$$

It follows that the required integral may be estimated from

$$\theta_4 = \frac{1}{N}\sum_{l=1}^{N}\frac{f(\xi_l)}{p(\xi_l)}, \qquad (2.4)$$

where $\xi_1, \xi_2, \ldots$ are the values of a random quantity ξ. In particular, when ξ is uniformly distributed in the interval (a, b) then $p(x) = \frac{1}{b-a}$ and (2.4) becomes identical with (2.1).

The variance of the function $\zeta^{(4)}$ is

$$\mathrm{D}\zeta^{(4)} = \int_a^b \frac{f^2(x)}{p(x)}\, dx - J^2.$$

The problem now arises as to how to select ξ in order to ensure that the variance will be a minimum.

The minimum $\mathrm{D}\zeta^{(4)}$ is achieved when the random quantity $\xi = \hat{\xi}$ is distributed in (a, b) with a density

$$\hat{p}(x) = \frac{|f(x)|}{\int_a^b |f(x)|\, dx}.$$

In this case,

$$\mathrm{D}\hat{\zeta}^{(4)} = \left[\int_a^b |f(x)|\, dx\right]^2 - J^2,$$

and if the function $f(x)$ does not change sign then $\mathrm{D}\hat{\zeta}^{(4)} = 0$.

In actual fact the use of the random quantity $\hat{\xi}$ is pointless since this would involve the evaluation of the integral $\int_a^b |f(x)|\, dx$, which is practically equivalent to the evaluation of J (when $p(x)$ does not change sign the two are exactly equivalent).

The practical conclusion is: the quantity ξ should be chosen so that the ratio of $f(x)$ to $|f(x)|$ varies as little as possible, i.e.,

$$\frac{p(x)}{|f(x)|} \approx \text{const.}$$

Example. Evaluate the integral

$$J = \int_0^1 e^x\, dx.$$

Since $e^x = 1 + x + \ldots$, let us consider a random quantity ξ with a density $p(x) = \frac{2}{3}(1 + x)$. According to (2.4) we have

$$J_4 = \frac{3}{2N} \sum_{i=1}^{N} \frac{e^{\xi_i}}{1+\xi_i}.$$

The variance of the averaged quantity is then given by

$$D\zeta^4 = \frac{3}{2} \int_0^1 \frac{e^{2x}\,dx}{1+x} - J^2 =$$
$$\frac{3}{2} e^{-2} [\mathrm{Ei}(4) - \mathrm{Ei}(2)] - (e-1)^2 = 0.0269.^{*}$$

The values of ξ may be obtained from the formula

$$\xi = \sqrt{1+3\gamma} - 1,$$

where γ is uniformly distributed in $(0, 1)$. Numerical details of this calculation are given in paragraph 11.

Remark. Suppose it is required to evaluate the integral

$$J = \int_a^b f(x)\, p(x)\, dx,$$

where $p(x)$ is the probability density of a random quantity ξ. Here also it is possible to use selective sampling: instead of $J = \mathrm{M} f(\xi)$ it is possible to estimate

$$J = \mathrm{M} \left\{ \frac{f(\eta)\, p(\eta)}{\tilde{p}(\eta)} \right\},$$

where η is a random quantity with probability density $\tilde{p}(x)$.

8. Group Sampling

This method is well known in statistics [33]. It is similar in principal to the method of selective sampling (more random points are selected in the "more important" regions). However, when the Monte Carlo method is used, selective sampling is more frequently employed.

Let us divide the interval (a, b) into m intervals (a_k, b_k) with lengths l_k $(k = 1, 2, \ldots, m$; Fig. 5):

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \ldots < b_{m-1} = a_m < b_m = b;$$
$$l_1 + l_2 + \ldots + l_m = b - a.$$

---

*Ei (x) is an exponential integral function. See Matematicheskii analiz [Mathematical Analysis] (functions, limits, series, continued fractions), Fizmatgiz, 1961, p. 379.

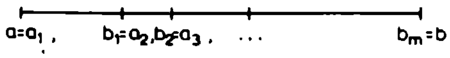


Fig. 5.

Each of the integrals following the summation sign in

$$\int_a^b f(x)\,dx=\sum_{k=1}^{m}\int_{a_k}^{b_k} f(x)\,dx$$

will be evaluated by the simple Monte Carlo method (2.1), using N_k values of quantity $\xi^{(k)}$ which is uniformly distributed in (a_k, b_k). This yields the following estimate

$$\theta_5=\sum_{k=1}^{m}\frac{l_k}{N_k}\sum_{l=1}^{N_k} f(\xi_l^{(k)}). \tag{2.5}$$

It is easy to show that

$$\mathrm{M}\theta_5=J, \quad \mathrm{D}\theta_5=\sum_{k=1}^{m}\frac{l_k^2 D_k}{N_k},$$

where $D_k=\mathrm{D}f(\xi^{(k)})$.*

The total number of values of $f(x)$ which are used in the calculations is

$$N=N_1+N_2+\dots+N_m.$$

It is easy to show that the minimum of the variance $\mathrm{D}\theta_5$ is reached when the $N_k=\hat{N}_k$ are proportional to $l_k\mathrm{D}_k^{\frac{1}{2}}$. In the latter case

$$\mathrm{D}\hat{\theta}_5=\frac{1}{N}\left(\sum_{k=1}^{m} l_k\sqrt{D_k}\right)^2.$$

However, the values of D_k are as a rule unknown at the beginning of the calculation. The quantities N_k are therefore frequently taken to be proportional to l_k:

---

*It is clear that

$$\mathrm{D}f(\xi^{(k)})=\frac{1}{b_k-a_k}\int_{a_k}^{b_k} f^2(x)\,dx-\left[\frac{1}{b_k-a_k}\int_{a_k}^{b_k} f(x)\,dx\right]^2.$$

$$\tilde{N}_k = \frac{l_k N}{b-a}.$$

In that case

$$\mathsf{D}\tilde{\theta}_5 = \frac{b-a}{N} \sum_{k=1}^{m} l_k D_k.$$

It is clear from the inequality

$$\mathsf{D}\hat{\theta}_5 \leqslant \mathsf{D}\tilde{\theta}_5 \leqslant \mathsf{D}\theta_1$$

that in both of the above two cases group sampling is more efficient than the estimate θ_1 (realization time $\tau_1 \approx \tau_5$).

It is obvious that the region of integration can be conveniently divided into parts in each of which the integrand varies slowly. In the multidimensional case it is often sensible to approximate the boundaries of these parts by some simple function $g(x)$ which has properties similar to the integrand.

Example. Evaluate the integral

$$J = \int_0^1 e^x\,dx.$$

Let us divide $(0, 1)$ into two equal parts and select four points in $(0, 1/2)$ and six points in $(1/2, 1)$. This yields

$$J_5 = \frac{1}{8} \sum_{i=1}^{4} e^{\xi_i^{(1)}} + \frac{1}{12} \sum_{i=1}^{6} e^{\xi_i^{(2)}}$$

and the variance is given by

$$\mathsf{D}J_5 = \frac{1}{16} D_1 + \frac{1}{24} D_2,$$

where

$$D_1 = 2\int_0^{1/2} e^{2x}\,dx - \left[2\int_0^{1/2} e^x\,dx\right]^2 = e - 1 - 4(\sqrt{e} - 1)^2 = 0.03492;$$

$$D_2 = 2\int_{1/2}^{1} e^{2x}\,dx - \left[2\int_{1/2}^{1} e^x\,dx\right]^2 = e^2 - e - 4(e - \sqrt{e})^2 = 0.09493.$$

It follows that $\mathrm{D}J_5 = 0.006138$ and $\delta_{\text{prob}} = 0.675(0.00614)^{1/2} = 0.053$. $\xi^{(1)}$ and $\xi^{(2)}$ may be calculated from the values of γ:

$$\xi^{(1)} = 0.5\gamma, \quad \xi^{(2)} = 0.5(1+\gamma).$$

Numerical details of this calculation are given in paragraph 11.

9. Symmetrization of the Integrand

Suppose it is required to evaluate

$$J = \int_0^1 f(x)\,dx.$$

We shall illustrate with examples how information about the behavior of $f(x)$ may be used to reduce the sampling variance.

Suppose that it is known that $f(x)$ is monotonic (or approximately monotonic). It may be expected that the symmetrized function

$$f_6(x) = \frac{1}{2}[f(x) + f(1-x)]$$

will vary more slowly than $f(x)$ (Fig. 6) and hence the variance of

$$\zeta^{(6)} = \frac{1}{2}[f(\gamma) + f(1-\gamma)]$$

will be smaller than $\mathrm{D}\zeta^{(1)}(\gamma)$. The corresponding estimate of the integral is

$$\theta_6 = \frac{1}{2N}\sum_{i=1}^{N}[f(\gamma_i) + f(1-\gamma_i)], \tag{2.6}$$

where $\gamma_1, \gamma_2, \ldots$ are the values of a random quantity γ which is uniformly distributed in $(0, 1)$. It is easy to show that

$$\mathrm{D}\zeta^{(6)} \leqslant \mathrm{D}\zeta^{(1)}.$$

However, it does not yet follow that (2.6) is always more convenient than (2.1). The time for the calculation using (2.6) is approximately twice as long as the time for (2.1), i.e.,

$$\tau_6 \approx 2\tau_1.$$

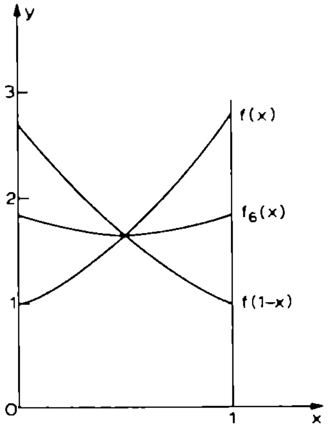


Fig. 6.

Hence (2.6) is more efficient than (2.1) provided

$$D\zeta^{(6)} \leqslant 0.5\, D\zeta^{(1)}$$

or

$$\int_0^1 f(x)\, f(1-x)\, dx \leqslant J^2.$$

Example. Evaluate the integral

$$J = \int_0^1 e^x\, dx.$$

From (2.6) we have

$$J_6 = \frac{1}{2N} \sum_{l=1}^{N} \left(e^{\gamma_l} + e^{1-\gamma_l}\right).$$

The variance of the averaged function (Fig. 6) is

$$D\zeta^{(6)} = \frac{1}{4}\,[2e + (e-1)(5-3e)] = 0.00392.$$

Numerical details of this calculation are given in paragraph 11.

Suppose now that the function $f(x)$ is known to have a single extremum in the neighborhood of the point $x = 0.5$. The use of (2.6) would only complicate the situation. However, it may be expected that the function

$$g(x)=\frac{1}{2}\left[f\left(\frac{x}{2}\right)+f\left(1-\frac{x}{2}\right)\right],$$

whose integral is equal to J as before, will be monotonic (Fig. 7) and can then be symmetrized in accordance with (2.6).

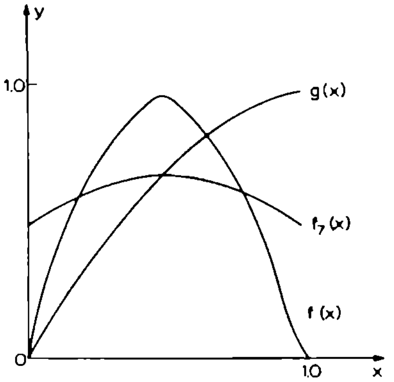


Fig. 7.

Thus

$$f_7(x)=\frac{1}{4}\left[f\left(\frac{x}{2}\right)+f\left(1-\frac{x}{2}\right)+f\left(\frac{1}{2}+\frac{x}{2}\right)+f\left(\frac{1}{2}-\frac{x}{2}\right)\right],$$

and the corresponding estimate of the integral J is

$$\theta_7=\frac{1}{4N}\sum_{i=1}^{N}\left[f\left(\frac{\gamma_i}{2}\right)+f\left(1-\frac{\gamma_i}{2}\right)+ f\left(\frac{1}{2}+\frac{\gamma_i}{2}\right)+f\left(\frac{1}{2}-\frac{\gamma_i}{2}\right)\right]. \tag{2.7}$$

This estimate is more efficient than (2.1) when $D\zeta^{(7)}\leqslant 0.25\,D\zeta^{(1)}$.

Example. Evaluate the integral

$$J=\int_0^1 \sin \pi x\,dx=\frac{2}{\pi}$$

with the aid of (2.7). Since in this example (Fig. 7)

$$f_7=\frac{\sqrt{2}}{2}\cos\pi\left(\frac{1}{4}-\frac{x}{2}\right),$$

it follows that the variance

$$D\zeta^{(7)}=\frac{1}{2}\int_0^1 \cos^2\pi\left(\frac{1}{4}-\frac{x}{2}\right)dx-\left(\frac{2}{\pi}\right)^2= \frac{1}{4}+\frac{1}{2\pi}-\frac{4}{\pi^2}=0.003871,$$

while

$$\mathsf{D}\zeta^{(1)}=\int_0^1 \sin^2 \pi x\, dx-\left(\frac{2}{\pi}\right)^2=\frac{1}{2}-\frac{4}{\pi^2}=0.09472,$$

so that $\mathsf{D}\zeta^{(1)}:\mathsf{D}\zeta^{(7)}=24.5$. Hence, in this case the estimate given by (2.7) is more efficient than that given by (2.1) by a factor of 6.1.

In contrast to the methods described in paragraphs 6-8, the application of symmetrization to multidimensional cases is not so convenient. It requires a preliminary transformation of the region of integration into a unit cube (corresponding to the number of dimensions).

In particular, in a three-dimensional unit cube the analog of the function

$$f_6(x)=\frac{1}{2}[f(x)+f(1-x)]$$

contains $2^3=8$ terms:

$$\begin{aligned}f_6(x, y, z)=\frac{1}{8}[&f(x, y, z)+f(1-x, y, z)+f(x, 1-y, z)+\\ &f(x, y, 1-z)+f(1-x, 1-y, z)+f(1-x, y, 1-z)+\\ &f(x, 1-y, 1-z)+f(1-x, 1-y, 1-z)].\end{aligned}$$

Some methods of symmetrization have been given by Morton [208] and Bakhvalov [3].

10. The Use of Dependent Quantities

This method is usually confined to estimates of the form of (2.3), (2.6) or (2.7), which may be looked upon as special cases of dependent quantities.

Suppose it is required to evaluate the integral

$$J=\int_a^b f(x)\, p_\xi(x)\, dx,$$

where $p_\xi(x)$ is the probability density of the random quantity ξ in the interval (a, b). In other words $\mathsf{M}f(\xi)=J$. Consider now a random quantity η and let $p_{\xi\eta}(x, y)$ be the probability density for the point (ξ, η) so that

$$p_\xi(x)=\int p_{\xi\eta}(x, y)\, dy, \quad p_\eta(y)=\int_a^b p_{\xi\eta}(x, y)\, dx$$

(the range of variation of η will not be indicated). It is not assumed that ξ and η are independent, so that $p_{\xi\eta}(x, y) \neq p_{\xi}(x)\, p_{\eta}(y)$. Consider now the random quantity

$$\zeta = f(\xi) - g(\eta),$$

where $g(y)$ is also arbitrary. If $I = \mathrm{M}\, g(\eta)$, then

$$\mathrm{M}\zeta = J - I.$$

Thus,

$$\mathrm{D}\zeta = \mathrm{D}f + \mathrm{D}g - 2r\sqrt{\mathrm{D}f \cdot \mathrm{D}g}, \tag{2.8}$$

where r is the correlation coefficient,

$$r = \frac{\mathrm{M}\{[f(\xi) - J][g(\eta) - I]\}}{\sqrt{\mathrm{D}f(\xi) \cdot \mathrm{D}g(\eta)}}.$$

It is known [23] that $|r| \leqslant 1$. When ξ and η are independent then $r = 0$.

A. Consider now η and $g(\eta)$ and suppose that the mathematical expectation I is known. Since

$$J = \mathrm{M}\zeta + I,$$

it is sufficient to estimate $\mathrm{M}\zeta$ in order to determine J. It is clear from (2.8) that $\mathrm{D}\zeta$ will be small for large positive values of r (the absolute minimum $\mathrm{D}\zeta = 0$ is reached when $\eta = \xi$ and $g = f$). It follows that the sampling of ζ is more convenient than the sampling of ξ when $g(\eta)$ and $f(\xi)$ have a large positive correlation. In particular, it is convenient when $\eta = \xi$ and $g(x) \approx f(x)$, as was the case in paragraph 6.

B. Let us now take η and $g(\eta)$ so that $I = -J$. Here an estimate of $\mathrm{M}\zeta$ will also yield J since

$$J = \frac{1}{2}\mathrm{M}\zeta.$$

Again, the use of ζ is more convenient than the use of ξ when $g(\eta)$ and $f(\xi)$ have a large positive correlation.

For example, if $p_{\xi}(x) = 1$, as was the case in paragraph 9, then we can have $\eta = \xi$ and $g(y) = -f(1 - y)$.

Further development of these methods has led to "random" quadrature formulas. In fact, the function

$$\zeta^{(7)} = \frac{1}{4}\left[f\left(\frac{\xi}{2}\right) + f\left(1 - \frac{\xi}{2}\right) + f\left(\frac{1}{2} + \frac{\xi}{2}\right) + f\left(\frac{1}{2} - \frac{\xi}{2}\right)\right]$$

which enters into (2.7) is a quadrature formula for each ξ. More complicated functions of the form

$$\zeta = \sum_{i=1}^{m} c_i f(h_i(\xi))$$

can also be used, where $h_i(\xi)$ are linear functions of ξ. Such estimates have been investigated by Hammersley and Morton [149] and Hammersley and Mauldon [147].* It may be noted that for sufficiently smooth functions the convergence with respect to m is more rapid than the convergence with respect to N, and it is convenient to use large m and small N (even $N = 1$).

Morton [208] and Bakhvalov [3] have developed analogous estimates for the multidimensional case. So far, they have not been of great practical importance because of their complexity [11].

Another approach to the construction of random quadrature formulas has been adopted by Ermakov and Zolotukhin [35].

11. Numerical Example

Various methods of evaluating the integral

$$J = \int_0^1 e^x \, dx = 1.718.$$

have been considered above. Here, this integral will be evaluated using all the above six methods with $N = 10$.** The values of the random quantity γ which is uniformly distributed in (0, 1) were taken from Table I of the Appendix. The values of random quantities with other distribution laws are obtained by transforming these values of γ.

The formulas for the evaluation of J are as follows

$$J_1 = \frac{1}{10} \sum_{i=1}^{10} e^{\gamma_i};$$

$$J_2 = \frac{e}{10} \sum_{i=1}^{10} g(\gamma_i, \gamma_{10+i}), \text{ where } g(x, y) = \begin{cases} 1 \text{ when } ey < e^x; \\ 0 \text{ when } ey \geqslant e^x; \end{cases}$$

---

*In particular, the methods A and B above are sometimes called the control variate and the antithetic variate methods, respectively.

**It is usually assumed in statistics that the central limit theorem holds for $N > 30$, so when $N = 10$ the estimate of the statistical error may be rather different from the true value. $N = 10$ was chosen in order to simplify the calculations.

$$J_3 = \frac{1}{2} + \frac{1}{10}\sum_{i=1}^{10}\left(e^{\gamma_i} - \gamma_i\right);$$

$$J_4 = \frac{3}{20}\sum_{i=1}^{10}\frac{e^{\xi_i}}{1+\xi_i}, \text{where } \xi_i = \sqrt{1+3\gamma_i} - 1;$$

$$J_5 = \frac{1}{8}\sum_{i=1}^{4} e^{\xi_i^{(1)}} + \frac{1}{12}\sum_{i=1}^{6} e^{\xi_i^{(2)}}, \text{where } \xi_i^{(1)} = \frac{\gamma_i}{2},\ \xi_i^{(2)} = \frac{1+\gamma_{4+i}}{2};$$

$$J_6 = \frac{1}{20}\sum_{i=1}^{10}\left(e^{\gamma_i} + e^{1-\gamma_i}\right).$$

The results of the calculations are given in Table 2. The values of the probable errors are also given for comparison. It is clear from the table that the actual errors are of the order of δ_{prob}.

Table 2

k	1	2	3	4	5	6
J_k	1.901	1.359	1.798	1.782	1.804	1.729
$J_k - J$	0.183	−0.359	0.080	0.064	0.086	0.011
δ_{prob}	0.10	0.28	0.044	0.035	0.053	0.013

Table 3 may be used to estimate the efficiency of all the estimates J_k in the case of the "Strela" computer (average speed of 3000 operations per second); τ_k is the time necessary to evaluate J_k with the aid of this computer in milliseconds.

Table 3

k	1	2	3	4	5	6
$\mathrm{D}J_k$	0.0242	0.1718	0.00437	0.00270	0.00614	0.000391
τ_k	153	179	157	402	161	295
$\tau_k \cdot \mathrm{D}J_k$	3.70	30.7	0.69	1.09	0.98	0.12

The progress of the calculations can easily be followed from Table 4, which also gives the intermediate quantities.

Table 4

i	γ_i	e^{γ_i}	γ_{10+i}	$g(\gamma_i;\gamma_{10+i})$	ξ_i	e^{ξ_i}	$e^{\xi_i}/(1+\xi_i)$	$\xi_i^{(1),(2)}$	$e^{\xi_i^{(1),(2)}}$	$e^{1-\gamma_i}$
1	0.86515	2.3754	0.42502	1	0.89617	2.4502	1.2922	0.43257	1.5412	1.1443
2	0.90795	2.4793	0.99224	0	0.92973	2.5338	1.3130	0.45398	1.5746	1.0964
3	0.66155	1.9378	0.88955	0	0.72761	2.0701	1.1982	0.33077	1.3920	1.4028
4	0.66434	1.9432	0.53758	1	0.73004	2.0752	1.1995	0.33217	1.3940	1.3989
5	0.56558	1.7604	0.91641	0	0.64217	1.9006	1.1574	0.78279	2.1875	1.5441
6	0.12332	1.1312	0.18867	1	0.17045	1.1858	1.0131	0.56166	1.7536	2.4029
7	0.94377	2.5696	0.41686	1	0.95737	2.6049	1.3308	0.97188	2.6429	1.0579
8	0.57802	1.7825	0.42163	1	0.65350	1.9223	1.1626	0.78901	2.2012	1.5250
9	0.69186	1.9974	0.85181	0	0.75373	2.1249	1.2116	0.84593	2.3301	1.3609
10	0.03393	1.0345	0.38967	0	0.04966	1.0509	1.0012	0.51697	1.6769	2.6276
Total	6.03547	19.0113		5			11.8796			15.5608

3. EVALUATION OF MULTIDIMENSIONAL INTEGRALS

12. Simple Methods

All the methods and estimates of Section 2 can easily be extended to multidimensional integrals of the form

$$\int_G f(P)\,dP, \tag{2.9}$$

where G is an arbitrary region of a d-dimensional space, the points $P=(x_1, \ldots, x_d)$ belong to the region G, and $dP = dx_1\,dx_2 \ldots dx_d$.

Example. Some of the integrals which are encountered in mathematical physics (potential theory, scattering theory, and so on) may be reduced to the form

$$J = \frac{1}{\pi^2} \int_S \int_S f(\rho)\,dP\,dQ,$$

where the points P and Q lie within the unit sphere S and ρ is the distance between them:

$$\rho = |P - Q|.$$

This is, in fact, a triple integral. However, it may be regarded as a sextuple integral and the symmetry may be taken into account during the realization of the random trials.

Let P and Q be two independent random points which are uniformly distributed in the sphere S, i.e., their densities inside S are given by

$$p(P) = p(Q) = \left(\frac{4}{3}\pi\right)^{-1}.$$

In terms of spherical coordinates (r, φ, μ), where $\mu = \cos\theta$ (Fig. 8), the volume element is given by $dP = r^2\,dr\,d\varphi\,d\mu$. In view of the symmetry of the problem the point P may be chosen to lie on the Oz axis so that

$$3 \int_0^{r_P} r^2\,dr = \gamma_1.$$

The point Q may be chosen to lie in the Oxz plane $(\varphi = 0)$ so that

$$\int_{-1}^{\mu_Q} \frac{d\mu}{2} = \gamma_2, \quad 3 \int_0^{r_Q} r^2\,dr = \gamma_3.$$

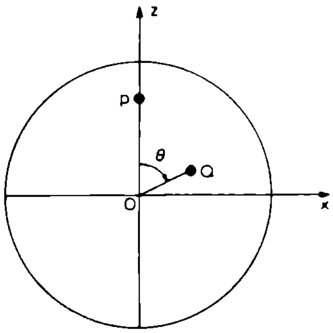


Fig. 8.

The resulting formulas used in the computation are then

$$r_P = \sqrt[3]{\gamma_1}, \quad \mu_Q = 2\gamma_2 - 1, \quad r_Q = \sqrt[3]{\gamma_3},$$
$$\rho = \sqrt{r_P^2 + r_Q^2 - 2\mu_Q r_P r_Q}.$$

The approximate value of the integral is

$$\bar{\theta} = \frac{(4/3)^2}{N} \sum_{i=1}^{N} f(\rho_i), \tag{2.10}$$

where ρ_i is the value of ρ obtained as a result of the ith trial.

13. Methods of Reducing the Variance

The following rule must be borne in mind in addition to the methods of reducing the variance which were given in Section 2: the variance is reduced by analytical integration with respect to one of the variables. Thus, suppose that it is required to evaluate the integral

$$J = \int_{V_P} \int_{V_Q} f(P, Q)\, p(P, Q)\, dP\, dQ,$$

where P and Q are points belonging to the spaces V_P and V_Q, respectively, and $p(P, Q)$ is the probability density. Suppose further that the integration with respect to Q may be carried out analytically, so that the partial probability density

$$\tilde{p}(P) = \int_{V_Q} p(P, Q)\, dQ$$

is known, and so is the function $\tilde{f}(P)$ which is defined by

$$\tilde{f}(P)\,\tilde{p}(P)=\int\limits_{V_Q} f(P,\ Q)\,p(P,\ Q)\,dQ.$$

It is clear that

$$J=\int\limits_{V_P}\tilde{f}(P)\,\tilde{p}(P)\,dP.$$

This means that the calculation of the average $\mathsf{M}f$ over the density $p(P,\ Q)$ may be replaced by the calculation of the average $\mathsf{M}\tilde{f}$ over the density $\tilde{p}(P)$. Hence

$$\mathsf{D}f-\mathsf{D}\tilde{f}=\int\limits_{V_P}\int\limits_{V_Q} f^2(P,\ Q)\,p(P,\ Q)\,dP\,dQ - \int\limits_{V_P}\tilde{f}^2(P)\,\tilde{p}(P)\,dP \geqslant 0.$$

This difference is equal to the average value of the variance $\mathsf{D}_p f$ which characterizes the spread of Q for a fixed P:

$$\mathsf{D}f-\mathsf{D}\tilde{f}=\int\limits_{V_P}\mathsf{D}_p f\tilde{p}(P)\,dP.$$

In particular, if the density $p(P,\ Q)$ is constant in $V_P \times V_Q$, then the density $\tilde{p}(P)$ is constant in V_P.

14. Integrals With Singularities

Suppose that the region G is finite and the function $f(P)$ has a singularity inside G or on the boundary. It is required to evaluate the integral (2.9):

$$\int\limits_{G} f(P)\,dP.$$

Here, it is recommended that selective sampling should be used, with the density $p(P)$ having the same singularity as the integrand $f(P)$.

The simple Monte Carlo method may be used to evaluate (2.9) only when $\int\limits_{G} f^2(P)\,dP$ converges; in the opposite case $\mathsf{D}\theta_1=\infty$. However, when the variance $\mathsf{D}\theta_1$ is finite, the inclusion of the singularity in the density usually leads to a reduction in the variance.

If the region G is not bounded, then the integral (2.9) may be evaluated as follows:

1) Parts of the integral which correspond to sufficiently distant parts of G are rejected, and the integral is evaluated over the remaining finite regions, or

2) The integral is transformed so that it corresponds to the case considered above, or

3) Selective sampling is used, with the probability density decreasing sufficiently rapidly at infinity.

The last-named method is clearly the most natural. The singularity can again be included in the density so that the probability density falls off at infinity at the same rate as the integrand.

Example. Suppose that in the integral

$$J=\frac{1}{\pi^2}\int\limits_S\int\limits_S f(\rho)\,dP\,dQ$$

(cf. paragraph 12 for notation), the integrand has a singularity of the form ρ^{-2}. $\bar{\theta}$ will then yield a poor estimate of the integral, since $D\bar{\theta}=\infty$.*

Suppose that the random point P is uniformly distributed in S so that $p(P)=3\,(4\pi)$. We shall try to find the random point Q which is such that its density is proportional to ρ^{-2}.

It is convenient to use spherical coordinates (ρ, φ, μ) with the origin at P (Fig. 9). Consider an arbitrary direction $\mu=\cos\theta$ in the Oxz plane $(\varphi=0)$. Let $l=l(r_P, \mu)$ be the distance of the point P from the surface of the sphere in this direction (PL in Fig. 9), and let ρ be uniformly distributed in the interval $(0, l)$. Under these conditions the density is given by**

$$p(Q)=\frac{1}{4\pi\rho^2 l}.$$

The computational formulas are then of the form

$$r_P=\sqrt[3]{\gamma_1},\quad \mu=2\gamma_2-1,$$
$$l=\mu r_P+\sqrt{1-r_P^2(1-\mu^2)},$$
$$\rho=\gamma_3 l.$$

---

*The integral $\int\limits_S\int\limits_S f^2(\rho)\,dP\,dQ\sim\int\limits_S\int\limits_S\frac{dP\,dQ}{\rho^4}$ converges.

**A more correct notation is $p\,(Q/P)$ since this is the conditional probability density.

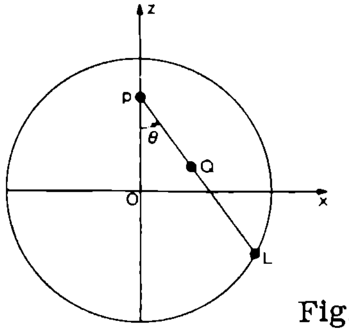


Fig. 9.

The approximate value of the integral is

$$\tilde{\theta} = \frac{{}^{16}/_{3}}{N} \sum_{i=1}^{N} \rho_i^2 f(\rho_i) l_i. \tag{2.11}$$

where ρ_i and l_i are the values of ρ and l obtained as a result of the ith trial.

15. Numerical Example

We shall use the estimates $\bar{\theta}$ and $\tilde{\theta}$ obtained in paragraphs 12 and 14 for the approximate evaluation of the integral

$$I_m = \frac{1}{\pi^2} \int_S \int_S \frac{dP\,dQ}{\rho^m},$$

for $m = 1$ and $m = 2$. In this integral S is a unit sphere in three-dimensional space, and the points P and Q lie within S; $\rho = |P - Q|$ is the distance between P and Q.

The integral I_m may be evaluated analytically.* In particular,

$$I_2 = 4, \quad I_1 = \frac{32}{15}, \quad I_{-1} = \frac{64}{35}.$$

A. Various estimates and their variances. Using (2.10) we have

$$\bar{I}_1 = \frac{({}^4/_3)^2}{N} \sum_{i=1}^{N} \frac{1}{\rho_i},$$

$$\bar{I}_2 = \frac{({}^4/_3)^2}{N} \sum_{i=1}^{N} \frac{1}{\rho_i^2}.$$

---

*When $m < 2$

$$I_m = \frac{32}{(2-m)(6-m)} \sum_{k=0}^{E\left(\frac{1-m}{2}\right)} \binom{2-m}{2k+1} \frac{1}{2k+3},$$

where $E(x)$ is the integral part of x.

The corresponding variances are

$$\mathsf{D}\bar{I}_1 = \frac{1}{N}\left[\left(\frac{4}{3}\right)^2 I_2 - I_1^2\right] = \frac{2.56}{N},$$
$$\mathsf{D}\bar{I}_2 = \infty.$$

Using (2.11) we have

$$\tilde{I}_1 = \frac{16/3}{N}\sum_{i=1}^{N} \rho_i l_i,$$
$$\tilde{I}_2 = \frac{16/3}{N}\sum_{i=1}^{N} l_i.$$

The corresponding variances are

$$\mathsf{D}\tilde{I}_1 = \frac{1}{N}\left[\frac{64}{9} I_{-1} - I_1^2\right] = \frac{8.452}{N},$$
$$\mathsf{D}\tilde{I}_2 = \frac{1}{N}\left[\frac{32}{3} I_1 - I_2^2\right] = \frac{6.756}{N}$$

More explicitly:

$$I_2 = \mathsf{M}\left\{\frac{16}{3} l\right\};$$

$$\mathsf{M}\left\{\left(\frac{16}{3}\right)^2 l^2\right\} = \left(\frac{16}{3}\right)^2 \int_0^1 3r^2\,dr \int_{-1}^{1} \frac{d\mu}{2}\, l^2 =$$

$$\frac{16^2}{3}\int_0^1 r^2\,dr \int_{-1}^{1} d\mu \int_0^l \rho\,d\rho = \frac{16^2}{3\cdot 2\pi}\int_0^1 r^2\,dr \int_S \frac{dQ}{\rho} =$$

$$\frac{16^2}{3}\,\frac{1}{8\pi^2}\int_S\int_S \frac{dP\,dQ}{\rho} = \frac{32}{3} I_1;$$

$$\mathsf{D}\tilde{I}_2 = \frac{1}{N}\,\mathsf{D}\left\{\frac{16}{3} l\right\} = \frac{1}{N}\left[\mathsf{M}\left\{\left(\frac{16}{3}\right)^2 l^2\right\} - I_2^2\right].$$

It is interesting to note that $\mathsf{D}\tilde{I}_1 > \mathsf{D}\bar{I}_1$, and hence it is clear that if the integrand $f(\rho)$ in the example of paragraph 14 does not have a singularity of the form ρ^{-2} then $\tilde{\theta}$ may be a worse estimate than $\bar{\theta}$. It follows that an incorrectly used selective sampling may yield an inferior result as compared with the simple estimate $\bar{\theta}$.

Next, it is easy to see that a statistical determination of ρ is not required for the evaluation of $\tilde{I}_2$. In other words, it may be considered that integration with respect to ρ in I_2 has been carried out analytically. An analogous estimate may be obtained for the integral I_1:

$$\hat{I}_1 = \frac{8/3}{N}\sum_{i=1}^{N} l_i^2.$$

In fact,

$$\mathsf{M}\hat{I}_1 = \frac{8}{3}\mathsf{M}\{l^2\} = \frac{8}{3}\int_0^1 3r^2\,dr \int_{-1}^{1}\frac{d\mu}{2}\,l^2 = 8\int_0^1 r^2\,dr\int_{-1}^{1} d\mu\int_0^l \rho\,d\rho = I_1,$$

and the variance of this estimate is

$$\mathsf{D}\hat{I}_1 = \frac{1}{N}\left[\frac{16}{3}I_{-1} - I_1^2\right] = \frac{5.201}{N}.$$

It is clear that $\mathsf{D}\hat{I}_1 < \mathsf{D}\tilde{I}_1$.

B. Calculations. The integrals I_1 and I_2 were evaluated by all the above methods for $N = 10$ (cf. footnote on page 58). The random quantities γ which are uniformly distributed in $(0, 1)$ were taken from Table I in the Appendix.

The formulas employed were as follows:

$$\begin{aligned}
r_P &= \sqrt[3]{\gamma_1}\\
\mu_Q &= 2\gamma_2 - 1\\
r_Q &= \sqrt[3]{\gamma_3}\\
\rho &= \sqrt{r_P^2 + r_Q^2 - 2\mu_Q r_P r_Q}\\
\bar{I}_1 &= \frac{16}{90}\sum_{i=1}^{10}\frac{1}{\rho_i}\\
\bar{I}_2 &= \frac{16}{90}\sum_{i=1}^{10}\frac{1}{\rho_i^2}
\end{aligned}$$

$$\begin{aligned}
r_P &= \sqrt[3]{\gamma_1}\\
\mu &= 2\gamma_2 - 1\\
l &= \mu r_P + \sqrt{1 - r_P^2(1 - \mu^2)}\\
\rho &= \gamma_3 l\\
\tilde{I}_1 &= \frac{16}{30}\sum_{i=1}^{10}\rho_i l_i\\
\tilde{I}_2 &= \frac{16}{30}\sum_{i=1}^{10} l_i\\
\hat{I}_1 &= \frac{8}{30}\sum_{i=1}^{10} l_i^2
\end{aligned}$$

The results are given in Table 5 together with the values of the probable errors δ_{prob}.

The intermediate quantities which indicate the progress of the calculations are given in Table 6.

Table 5

	$\bar{I}_1$	$\tilde{I}_1$	$\hat{I}_1$	$\bar{I}_2$	$\tilde{I}_2$
Result	1.78	1.67	1.48	1.94	3.37
Error	−0.35	−0.46	−0.65	−2.06	−0.63
δ_{prob}	0.34	0.62	0.49	∞	0.55

C. The values of Table 6 may be used to obtain an empirical estimate of the variances of $\bar{I}_1$ and (cf. paragraph 2):

$$D\bar{I}_1 \approx \frac{1}{9}\left[\frac{1}{10}\left(\frac{16}{9}\right)^2 \sum_{i=1}^{10} \frac{1}{\rho_i^2} - \bar{I}_1^2\right] = 0.031,$$

$$D\bar{I}_2 \approx \frac{1}{9}\left[\frac{1}{10}\left(\frac{16}{3}\right)^2 \sum_{i=1}^{10} t_i^2 - \bar{I}_2^2\right] = 0.49.$$

The true values are respectively 0.26 and 0.68. The accuracy of the empirical value of $D\bar{I}_2$ is quite sufficient for the estimation of δ_{prob}. However, the empirical value of $D\bar{I}_1$ is not satisfactory because of the low accuracy of $\bar{I}_2$.

16. Monte Carlo Method With Increased Rate of Convergence

Consider the evaluation of the integral

$$J = \int_K f(P)\,dP, \tag{2.12}$$

where K is a unit cube in the d-dimensional space $0 \leqslant x_1, x_2, \ldots, x_d \leqslant 1$, the point $P = (x_1, \ldots, x_d)$ belongs to K and $dP = dx_1\,dx_2 \ldots dx_d$. The simple Monte Carlo estimation given by (2.1) can be used to evaluate (2.12) only if the integrand $f(P)$ has an integrable square. The order of convergence is then $N^{-\frac{1}{2}}$.

Quadrature formulas cannot be constructed for such a wide class of functions (cf. paragraph 17). However, quadrature formulas giving a higher rate of convergence are available for more restricted classes of functions. The known properties of the functions in a particular class are used to construct such formulas, and it is natural to expect that the same information may be used to develop Monte Carlo methods giving a more rapid rate of convergence.

Table 6

i	γ_1	γ_2	γ_3	r_P	$\mu_Q = \mu$	r_Q	$\frac{1}{\rho}$	$\frac{1}{\rho^2}$	l	l^2	$\gamma_0 l^2$
1	0.86515	0.90795	0.66155	0.9529	0.8159	0.8713	1.789	3.201	1.612	2.598	1.719
2	0.66434	0.56558	0.12332	0.8726	0.1312	0.4978	1.057	1.117	0.616	0.380	0.047
3	0.94377	0.57802	0.69186	0.9809	0.1560	0.8844	0.824	0.679	0.401	0.160	0.111
4	0.03393	0.42502	0.99224	0.3237	—0.1500	0.9974	0.914	0.836	0.899	0.808	0.802
5	0.88955	0.53758	0.91641	0.9617	—0.0752	0.9713	0.761	0.579	0.356	0.127	0.116
6	0.18867	0.41686	0.42163	0.5735	—0.1663	0.7499	0.983	0.967	0.729	0.532	0.224
7	0.85181	0.38967	0.33181	0.9479	—0.2207	0.6923	0.775	0.600	0.172	0.030	0.010
8	0.72664	0.53807	0.00607	0.8990	0.0761	0.1824	1.107	1.225	0.512	0.262	0.002
9	0.86522	0.47171	0.88059	0.9529	—0.0566	0.9585	0.720	0.518	0.254	0.065	0.057
10	0.89342	0.67248	0.09082	0.9631	0.3450	0.4495	1.097	1.204	0.760	0.577	0.052
Total							10.027	10.926	6.311	5.539	3.140

Consider a set of functions for which all the first order partial derivatives

$$\frac{\partial f}{\partial x_1},\ \frac{\partial f}{\partial x_2},\ \ldots,\ \frac{\partial f}{\partial x_d}$$

are continuous and bounded in K. There are quadrature formulas which ensure convergence of the order of $N^{-\frac{1}{d}}$ for this class of functions. There are also Monte Carlo methods which ensure convergence of the order of $N^{-\left(\frac{1}{d}+\frac{1}{2}\right)}$ for the same class of functions.* Bakhvalov [3] has shown that these rates of convergence cannot be improved upon. The difference between them is due to the fact that in the case of the quadrature formulas it is the maximum error for the particular class of functions which is being estimated, while in the case of the Monte Carlo method an estimate is made of the probability of the error for one of the functions in the class.

As an example of a quadrature formula ensuring convergence of the order of $N^{-\frac{1}{d}}$ we may take the formula

$$J \approx \frac{1}{N} \sum_{i=1}^{N} f(P_i), \tag{2.13}$$

in which the nodes P_i form a uniform (cubic) lattice in K. This means that the cube K is divided into $N = n^d$ equal cubes K_i (Fig. 10) and the point P_i is taken at the center of K_i.

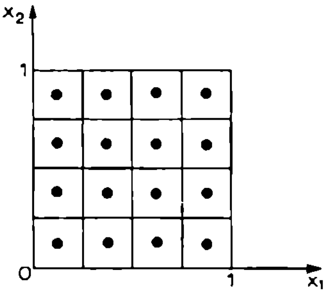


Fig. 10.

Consider now the formula given by (2.13), but suppose that the point P_i is a random point which is uniformly distributed in K_i. It is easy to show that

---

*In each case the integral is estimated using values of the function at N points.

$$\mathsf{M}\left\{\frac{1}{N}\sum_{i=1}^{N} f(P_i)\right\} = J;$$

$$\mathsf{D}\left\{\frac{1}{N}\sum_{i=1}^{N} f(P_i)\right\} \leqslant \frac{A^2}{n^{d+2}},$$

where the constant A depends only on d and the maximum of the moduli of the partial derivatives* $\left|\frac{\partial f}{\partial x_j}\right|$.

It follows from Chebyshev's inequality that

$$\mathsf{P}\left\{\left|\frac{1}{N}\sum_{j=1}^{N} f(P_j) - J\right| \leqslant \frac{A/\varepsilon}{N^{\frac{1}{2}+\frac{1}{d}}}\right\} \geqslant 1 - \varepsilon^2.$$

Hence, in this case the order of the convergence is $N^{-\left(\frac{1}{2}+\frac{1}{d}\right)}$.

For large d this method of calculation is hardly appropriate, since $\frac{1}{2}+\frac{1}{d}\approx\frac{1}{2}$. It is better to use the simple formula (2.1).

The Monte Carlo method involving increased rate of convergence has not been widely used in computational practice. It is probable that this is due to the fact that the multidimensional quadrature formulas which should serve as the basis for such Monte Carlo methods are very complicated.

17. Practical Remarks

A. Some properties of the Monte Carlo method. The first property is the probabilistic nature of the convergence. Although the order of convergence, $N^{-\frac{1}{2}}$, is not high, it is independent of the order d of the integral (cf. B, below). Moreover, the probable error of the calculations may be estimated in the course of the calculations themselves. If the result is required to a low accuracy (of the order of 5%) then this convergence is quite acceptable.

The second property of the method is its simplicity, which facilitates programming fast computers.

B. On the application of uniform nets. Uniform nets are frequently used in computational practice to evaluate multidimensional integrals over a unit cube K. Such nets are simple, and are simple generalizations of the procedure used with single (one-dimensional) integrals. However, uniform nets are not suitable in the case of

---

$*A = \sqrt{\frac{7}{60}}\,Ld$, where $L = \max_j \sup_K \left|\frac{\partial f}{\partial x_j}\right|$.

large d [56]. At first sight this would seem to contradict the statements made in paragraph 16, where it was shown that uniform nets ensure the best order of convergence $\left(N^{-\frac{1}{d}}\right)$ for the class of functions having continuous and bounded first-order partial derivatives. It is possible, however, to restrict the class of functions somewhat, and require that all the partial derivatives containing no more than one differentiation with respect to each variable should be continuous and bounded. In other words, the derivative $\frac{\partial^d f}{\partial x_1 \partial x_2 \ldots \partial x_d}$ and all the derivatives from which it can be obtained by differentiation should be continuous and bounded. For this class of functions it is possible to construct a net in the cube K which will ensure convergence of the order of $\frac{\ln^{d-1} N}{N}$, while the order of the convergence for uniform nets is $N^{-\frac{1}{d}}$, as before. For large d the difference is very considerable.

C. On the application of the Monte Carlo method. Consider the following three cases:

1) The function $f(P)$ is "sufficiently smooth" and the region G is "sufficiently good."* Such integrals are, in practice, evaluated by the Monte Carlo method for $d > 4$ and occasionally for $d \geqslant 3$. However, this is apparently a temporary situation: it is possible to construct quadrature formulas for various classes of such functions which have a faster rate of convergence. When such sufficiently simple and universal formulas are found, the Monte Carlo methods will become less convenient (cf. Section 5).

2) The function $f(P)$ is "piecewise smooth."** The integrals of such functions are evaluated by the Monte Carlo method for $d \geqslant 2$. In fact, in the case of a function of a single variable, the region in which it is defined may be divided into a number of intervals in each of which a suitable quadrature formula may be used. However, when $d = 2$ the division of a square into curvilinear figures and the transformation of each of them into a standard region for which adequate quadrature formulas are available is a complicated matter which is very inconvenient in computer practice. On the other hand, the computational process in the case of the simple Monte Carlo method does not depend on the position of the lines

---

*It may be imagined that G may be easily transformed into a unit d-dimensional cube K and the function $f(P)$ has continuous partial derivatives of order up to and including d.

**This includes the case of a "poor" bounded region G since it can always be included in a cube by adding a null function.

of discontinuity. The Monte Carlo method has, so far, been the most convenient method for such integrals.

3) The function $f(P)$ is "very poor." There are broad classes of functions for which quadrature formulas with convergence of order better than $N^{-\frac{1}{2}}$ cannot be constructed. In the one-dimensional case, functions of this class are, for example, those satisfying the Lipschitz condition of order α for $\alpha < \frac{1}{2}$:

$$|f(x) - f(x')| \leqslant L|x' - x|^{\alpha}.$$

Analogous classes of functions H_{α} for functions of a number of variables are given by Sobol' [60].

If the only thing which is known about a function $f(P)$ is that it belongs to a class which is broader than H_{α} for $\alpha < \frac{1}{2}$ (in this sense it is "very poor"), then as a rule its integral is best evaluated by the Monte Carlo method (even for $d = 1$).

4. EVALUATION OF WIENER INTEGRALS

Many important problems in quantum physics and theory of probability can be reduced to the evaluation of Wiener* integrals

$$\int\limits_C F[x]\, d_W x, \tag{2.14}$$

where C is the functional space of all functions $x = x(t)$ which are continuous in $[0, T]$ and satisfy the condition $x(0)=0$, and $F[x]$ is an arbitrary (continuous and bounded) functional defined in C.

18. Two Methods of Evaluating Wiener Integrals

The integral (2.14) may be evaluated approximately by replacing it by an integral of finite dimensions and sufficiently high multiplicity d. In fact, let us divide $[0, T]$ into d equal parts by the points

$$0 = t_0 < t_1 < t_2 < \ldots < t_d = T. \tag{2.15}$$

---

*We shall consider only the integral of Wiener measure. In determining this measure, the derivative should be made equal to $^1/_4$, which can always be arranged by changing the time scale. In this way, the probability density of $x(t)$ is

$$p_t(x) = \frac{1}{\sqrt{\pi t}} e^{-(x^2/t)}.$$

Each curve $x(t)$ will be replaced by a broken line $\tilde{x}(t)$ which intersects $x(t)$ at all the dividing points:

$$\tilde{x}(t_i) = x(t_i) = x_i.$$

The value of the functional $F[x]$ on such broken lines can be looked upon as a function of d variables:

$$F[x] = F(x_1, \ldots, x_d).$$

Hence,

$$\int_C F[x]\, d_W x =$$

$$\lim_{d \to \infty} \left(\frac{d}{\pi T}\right)^{\frac{d}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(x_1, \ldots, x_d)\, e^{-\frac{d}{T}\sum_{i=1}^{d}(x_i - x_{i-1})^2}\, dx_1 \ldots dx_d.$$

The multidimensional integrals on the right hand side may be evaluated by any method, including the Monte Carlo method.

A somewhat different approach is also possible. Wiener's measure in the space C defines the probability density for the trajectories of a Brownian particle satisfying the initial condition $x(0)=0$. Hence, the integral (2.14) represents the mathematical expectation (mean value) of the functional $F[x]$ over the complete set of such trajectories. For the approximate evaluation of (2.14) by the Monte Carlo method it is necessary to realize a sufficiently large number N of Brownian trajectories $x^{(1)}(t)$, $x^{(2)}(t)$, $\ldots$, $x^{(N)}(t)$, find the value of the functional for each of them, and average the result:

$$\int_C F[x]\, d_W x \approx \frac{1}{N} \sum_{k=1}^{N} F[x^{(k)}].$$

The realization of Brownian trajectories can also be approximately carried out by replacing them by broken lines.

19. Approximate Construction of Brownian Trajectories

In all the formulas given above, the quantities $\zeta_1, \zeta_2, \ldots$ are normal random quantities with a mean of zero and a variance of 1.

The first method is based directly on the definition of the Brownian motion. The interval $[0, T]$ is divided into d equal parts by the points defined by (2.15). The values of $x(t_i)$ are determined statistically using the formula

$$x(t_i)=x(t_{i-1})+\sqrt{\frac{T}{2d}}\zeta_i, \qquad i=1,\ 2,\ \ldots,\ d,$$

and the points of the trajectories, which are obtained as a result, are connected by straight lines.

The second method [20] is based on the fact that if the values of $x(s_1)$ and $x(s_2)$ are known, then the conditional distribution of $x\left(\frac{s_1+s_2}{2}\right)$ is normal with a mean equal to $\frac{1}{2}[x(s_1)+x(s_2)]$ and a variance of $\frac{1}{8}|s_2-s_1|$.

The computational formulas are

$$\begin{aligned}
&x(T)=\sqrt{\frac{T}{2}}\zeta_1,\\
&x\left(\frac{T}{2}\right)=\frac{1}{2}[x(0)+x(T)]+\sqrt{\frac{T}{8}}\zeta_2,\\
&x\left(\frac{T}{4}\right)=\frac{1}{2}\left[x(0)+x\left(\frac{T}{2}\right)\right]+\sqrt{\frac{T}{16}}\zeta_3,\\
&x\left(\frac{3T}{4}\right)=\frac{1}{2}\left[x\left(\frac{T}{2}\right)+x(T)\right]+\sqrt{\frac{T}{16}}\zeta_4,\\
&x\left(\frac{T}{8}\right)=\frac{1}{2}\left[x(0)+x\left(\frac{T}{4}\right)\right]+\sqrt{\frac{T}{32}}\zeta_5,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

Having determined 2^m points of the trajectory, the points are connected by straight lines.

20. Numerical Example

Evaluate the integral

$$\int_C \|x\|^2 d_W x=\frac{1}{4},$$

where

$$\|x\|^2=\int_0^1 x^2(t)\,dt.$$

Since

$$\mathrm{D}\{\|x\|^2\}=\int_C \|x\|^4 d_W x-\left(\frac{1}{4}\right)^2=\frac{7}{48}-\frac{1}{16}=\frac{1}{12},$$

it follows that for $N = 10$ the probable error is

$$\delta_{\text{prob}} = 0.675(0.00833) = 0.062$$

(in addition to this statistical error, there is also the further error due to the replacement of continuous trajectories by the broken lines).

The Brownian trajectories will be constructed by the second method using four points and the normal random quantities given in Table II of the Appendix. The number of trajectories is $N = 10$ and the computational formulas are

$$x(1) = 0.70711\zeta_1;$$

$$x(^1/_2) = 0.5x(1) + 0.35355\zeta_2;$$

$$x(^1/_4) = 0.5x(^1/_2) + 0.25\zeta_3;$$

$$x(^3/_4) = 0.5[x(^1/_2) + x(1)] + 0.25\zeta_4;$$

$$\|x\|^2 = \frac{1}{12}\{2x^2(^1/_4) + 2x^2(^1/_2) + 2x^2(^3/_4) + x^2(1) + x(^1/_4)\,x(^1/_2) + x(^1/_2)\,x(^3/_4) + x(^3/_4)\,x(1)\};$$

$$\int_C \|x\|^2 d_W x \approx \frac{1}{10}\sum_{i=1}^{10} \|x\|_i^2 *.$$

The results of the calculations are given in Table 7. The first three trajectories are shown in Fig. 11.

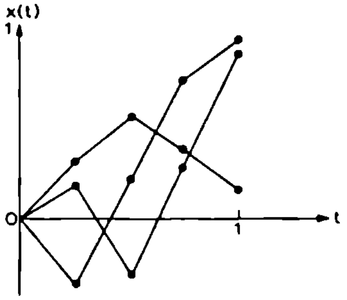


Fig. 11.

The final result (0.2585) turned out to be exceptionally good (by chance). For $N = 9$ the result is 0.2075, for which the error is of the order of δ_{prob}.

---

*The following formula was used: If the function $f(x)$ is linear in (a, b) then

$$\int_a^b f^2(x)\,dx = \frac{b-a}{3}[f^2(a) + f(a)f(b) + f^2(b)].$$

Table 7

l	ζ_1	ζ_2	ζ_3	ζ_4	$x(\frac{1}{4})$	$x(\frac{1}{2})$	$x(\frac{3}{4})$	$x(1)$	$\|x\|^2$	$\|x\|^4$
1	0.2005	1.1922	−0.0077	0.0348	0.2443	0.4924	0.3258	0.1418	0.0970	0.0094
2	1.0423	−1.8149	1.1803	0.0033	0.1585	−0.2732	0.2327	0.7370	0.0763	0.0058
3	1.1609	−0.6690	−1.5893	0.5816	−0.3104	0.1739	0.6428	0.8209	0.1949	0.0380
4	1.8818	0.7390	−0.2736	1.0828	0.3949	0.9266	1.3993	1.3306	0.9367	0.8774
5	0.5864	−0.9245	0.0904	1.5068	−0.0372	−0.1196	0.5242	0.4146	0.0760	0.0058
6	−1.1147	0.2776	0.1012	−1.3566	−0.1227	−0.2960	−0.8812	−0.7882	0.2809	0.0079
7	0.1425	−0.2863	1.2809	0.4043	0.2948	−0.0508	0.1261	0.1008	0.0177	0.0003
8	0.6379	−0.4428	−2.3006	−0.6446	−0.5406	0.0690	0.0989	0.4511	0.0693	0.0048
9	0.9516	−1.7708	2.8854	0.4686	0.5766	−0.2896	0.3088	0.6729	0.1190	0.0142
10	1.4664	1.6852	−0.9690	−0.0831	0.3149	1.1142	1.0548	1.0369	0.7168	0.5138
Total									2.5846	1.4774

The variance can easily be estimated from data given in Table 7:

$$\mathrm{D}\{\|x\|^2\} \approx \frac{1}{9}(1.4774 - 10 \cdot 0.2585^2) = 0.090$$

The exact value is $\frac{1}{12} = 0.083$.

5. APPLICATION OF QUASI-RANDOM POINTS TO THE MONTE CARLO SCHEME

It was noted in Section 3 (paragraph 17) that for "good" functions there exist multidimensional quadrature formulas ensuring a better rate of convergence than the Monte Carlo method. A simple method for the construction of such formulas will now be given [141, 60]. This method can be generalized to an infinite number of dimensions [61] and preserves the Monte Carlo scheme of calculations, except that the random quantities are replaced by quasi-random quantities.

Before we explain the method we must define the number $p_r(i)$.

21. The Numbers $p_r(i)$

Consider a positive integer $r \geqslant 2$. We shall give three equivalent definitions of $p_r(i)$.

Definition 1. If in the number system with scale of notation r

$$i = a_m a_{m-1} \ldots a_1 a_0,$$

then (in the same system)

$$p_r(i) = 0,\, a_0\, a_1 \ldots a_{m-1} a_m.$$

The quantities a_s are the integers $0 \leqslant a_s \leqslant r-1$. More explicitly: if

$$i = a_m r^m + a_{m-1} r^{m-1} + \ldots + a_1 r + a_0,$$

then

$$p_r(i) = \frac{a_0}{r} + \frac{a_1}{r^2} + \ldots + \frac{a_m}{r^{m+1}}.$$

For example, let $r = 2$, $i = 11$. Since in the binary system $i = 1011$, it follows that $p_r(i) = 0.1101$, or in the decimal system $p_2(11) = \frac{13}{16}$.

Definition 2 (recurrent).

a) $p_r(1) = \frac{1}{r}$;

b) If in the number system with scale of notation r

$$p_r(i) = 0,\ a_0 a_1 \ldots a_{m-1} a_m 00 \ldots,$$

then in order to obtain $p_r(i+1)$ it is necessary to find the smallest number k which is such that $a_k < r-1$; a_k is then replaced by $1+a_k$ and all the figures with smaller numbers (if they exist) are replaced by zeros; figures with numbers greater than k remain unaltered.

This rule may be written in the form

$$p_r(i+1) = p_r(i) + r^{-k} + r^{-k-1} - 1.$$

For example, in the binary system $p_2(11) = 0.1101$. In this case, $k = 2$. Hence, $p_2(12) = 0.0011$ or $p_2(12) = 3/16$.

Definition 3 (recurrent over groups).

a) $p_r(r^n) \quad = r^{-n-1}$ when $n = 0, 1, 2, \ldots$;

b) $p_r(r^n + j) = p_r(r^n) + p_r(j)$ when

$$j = 1, 2, \ldots, r^{n+1} - r^n - 1.$$

All the numbers $p_r(i)$ are rational and lie between zero and unity. For given r the sequence of points $p_r(1), p_r(2), \ldots, p_r(i)$ is uniformly distributed in the interval (0, 1) (in the number-theoretical sense).*

Some numerical values are given in Table 8.

Table 8

i	1	2	3	4	5	6	7	8
$p_2(i)$	$1/2$	$1/4$	$3/4$	$1/8$	$5/8$	$3/8$	$7/8$	$1/16$
$p_3(i)$	$1/3$	$2/3$	$1/9$	$4/9$	$7/9$	$2/9$	$5/9$	$8/9$
$p_5(i)$	$1/5$	$2/5$	$3/5$	$4/5$	$1/25$	$6/25$	$11/25$	$16/25$

22. Evaluation of Multidimensional Integrals

Suppose that K is a unit d-dimensional cube as before, $P = (x_1, x_2, \ldots, x_d)$ is a point in the cube, and $dP = dx_1\, dx_2 \ldots dx_d$. Consider d relatively prime integers $r_1, r_2, \ldots, r_d$.

---

*The sequence $p_2(1), p_2(2), \ldots, p_2(i), \ldots,$ was apparently first given by van der Korput and independently by Sobol' [56]. Generalization to arbitrary r is due to Hammersley.

Halton's sequence will be defined as the sequence of points $P_1^*, P_2^*, \ldots, P_l^*, \ldots$ in the cube K whose coordinates are given by

$$P_l^* = \left(p_{r_1}(l), \ p_{r_2}(l), \ \ldots, \ p_{r_d}(l)\right).$$

This sequence is uniformly distributed in the cube K (in the number-theoretical sense).

In the theorem given below, an estimate is made of the error in the quadrature formula

$$\int_K f(P)\,dP \approx \frac{1}{N}\sum_{l=1}^{N} f(P_l^*).$$

Definition. A class $W_1(L)$ of functions is defined as the set of functions $f(P)$ whose partial derivatives containing no more than one differentiation with respect to each of the coordinates* are piecewise continuous and bounded, so that

$$\left|\frac{\partial^m f(P)}{\partial x_{l_1}\,\partial x_{l_2}\ldots\partial x_{l_m}}\right| \leqslant L,$$

where $1 \leqslant l_1 < l_2 < \ldots < l_m \leqslant d$; $m = 1, 2, \ldots, d$.

Theorem. There exists a constant B (which depends on $r_1, r_2, \ldots, r_d$) which is such that for any $f(P)$ belonging to the class $W_1(L)$ the inequality

$$\left|\int_K f(P)\,dP - \frac{1}{N}\sum_{l=1}^{N} f(P_l^*)\right| \leqslant LB\frac{\ln^d N}{N} \qquad (2.16)$$

will hold for all integral N.

So far, it is not clear what is the best method of choosing $r_1, r_2, \ldots, r_d$. Known estimates of B merely indicate that small constants are the most convenient.

For sufficiently large N,

$$B = \prod_{l=1}^{d} \frac{r_l - 1}{\ln r_l}.$$

may be used in (2.16).

---

*Among the derivatives covered by this definition the highest order derivative is $\dfrac{\partial^d f}{\partial x_1\,\partial x_2\ldots\partial x_d}$.

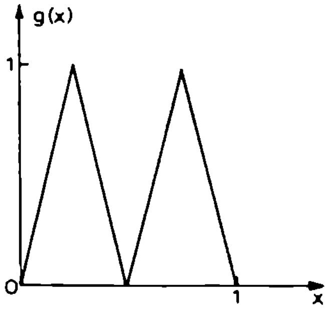


Fig. 12.

Example. Evaluate the integral

$$J = \int_0^1 \int_0^1 \int_0^1 f(x, y, z)\, dx\, dy\, dz,$$

where $f = g(x) g(y) g(z)$ and the function $g(x)$ (Fig. 12) is defined by

$$g(x) = \begin{cases} 4x, & 0 \leqslant x \leqslant \frac{1}{4}; \\ 2 - 4x, & \frac{1}{4} \leqslant x \leqslant \frac{1}{2}; \\ 4x - 2, & \frac{1}{2} \leqslant x \leqslant \frac{3}{4}; \\ 4 - 4x, & \frac{3}{4} \leqslant x \leqslant 1. \end{cases}$$

The exact value of the integral is

$$J = \left[\int_0^1 g(x)\, dx \right]^3 = 0.5^3 = 0.125.$$

Let r_1, r_2, r_3 be equal to 2, 3, 5 and let us estimate J from the eight points P_1^*, P_2^*, ..., P_8^*. Using the values of $p_r(i)$ given in paragraph 21 we obtain the values of $g(p_r(i))$ which are given in Table 9.

Table 9

i	1	2	3	4	5	6	7	8
$g(p_2(i))$	0	1	1	$1/2$	$1/2$	$1/2$	$1/2$	$1/4$
$g(p_3(i))$	$2/3$	$2/3$	$4/9$	$2/9$	$8/9$	$8/9$	$2/9$	$4/9$
$g(p_5(i))$	$4/5$	$2/5$	$2/5$	$4/5$	$4/25$	$24/25$	$6/25$	$14/25$

The approximate value of J is

$$\frac{1}{8}\sum_{i=1}^{8} g(p_2(i))\,g(p_3(i))\,g(p_5(i)) = 0.125 \cdot 1.120 = 0.140.$$

23. Evaluation of Infinitely Dimensional Integrals

In this paragraph K will denote the infinitely dimensional unit cube

$$K = \{0 \leqslant x_1 \leqslant 1,\ 0 \leqslant x_2 \leqslant 1,\ \ldots,\ 0 \leqslant x_s \leqslant 1,\ \ldots\}.$$

A point P of this cube is defined by the denumerable sequence of coordinates

$$P = (x_1,\ x_2,\ \ldots,\ x_s,\ \ldots),$$

and $dP = dx_1 dx_2 \ldots dx_s \ldots$ is an infinite product. Let us enumerate all the prime numbers: $r_1 < r_2 < \ldots < r_s < \ldots$ (i.e., $r_1 = 2$, $r_2 = 3$, $r_3 = 5$, $r_4 = 7, \ldots$).

The generalized Halton's sequence is defined as the sequence of points P_1^*, P_2^*, ..., P_i^*, ... of the cube K whose coordinates are defined by

$$P_i^* = \left(p_{r_1}(i),\ p_{r_2}(i),\ \ldots,\ p_{r_s}(i),\ \ldots\right).$$

This sequence may be used to evaluate certain infinitely dimensional integrals. Without giving a precise definition of the class of functions, we shall merely note that the piecewise continuity and boundedness of all the partial derivatives containing not more than one differentiation with respect to each coordinate are not sufficient.* A further necessary condition is that the dependence of $f(P)$ on x_s must decrease sufficiently rapidly with increasing s. This was first pointed out by N. N. Chentsov.

For such functions it may be shown that

$$\left| \int_K f(P)\,dP - \frac{1}{N}\sum_{i=1}^{N} f(P_i^*) \right| = O\left(\frac{1}{N^{1-\varepsilon}}\right),$$

where $\varepsilon > 0$ is a constant which is as small as desired.

---

*In this requirement are included the derivatives of any order.

24. The Points P_i^* as Deterministic Quasi-Random Points

In computational Monte Carlo practice the realization of the values of any random quantities is usually carried out by transforming the values of a random quantity γ which is uniformly distributed in the interval (0, 1) (cf. Chapter IV). Thus, instead of calculating $f(\xi_1, \xi_2, \ldots, \xi_d)$, where $(\xi_1, \xi_2, \ldots, \xi_d)$ is a random point having a given distribution law, one evaluates some other function

$$f(\xi_1, \xi_2, \ldots, \xi_d) = F(\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(n)}),$$

in which $\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(n)}$ are similar random quantities which are uniformly distributed in the interval (0, 1). Generally speaking, $n \geqslant d$.

Thus, the calculation of the average value of f in a given d-dimensional region is replaced by the calculation of the average value of the function F over an n-dimensional unit cube. The point

$$\Gamma = (\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(n)})$$

is a random point which is uniformly distributed in this cube.

According to paragraph 22, the average value of the function F over the n-dimensional unit cube may be calculated over the points of the Halton sequence. Hence, in some calculations the sequence of random points $\Gamma_1, \Gamma_2, \ldots$, which are uniformly distributed in the unit cube, may be replaced by $P_1^*, P_2^*, \ldots$. If the function F obtained as a result of this procedure turns out to be sufficiently smooth, then the method may ensure a more rapid convergence, namely, $\frac{\ln^n N}{N}$ instead of $N^{-\frac{1}{2}}$.

The solution of a number of physical problems by the Monte Carlo method involves the simulation of elementary particle trajectories (cf. Chapter III). Theoretically, each of these trajectories is defined by a denumerable sequence of values

$$\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(s)}, \ldots$$

and may be looked upon as a point of an infinitely dimensional cube, although in practice the calculation of the trajectory is cut off. The calculated quantities are functions of the trajectory; they are calculated for each trajectory and are then averaged over all trajectories.

It is clear that instead of the random, infinitely dimensional points $(\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(s)}, \ldots)$ it is occasionally possible to use the points of the generalized Halton sequence. This means that the numbers $p_{r_1}(i), p_{r_2}(i), \ldots, p_{r_s}(i), \ldots$ may be used instead of the quasi-random numbers $\gamma_1, \gamma_2, \ldots, \gamma_s, \ldots$ to construct the ith trajectory. It follows from paragraph 23 that in some problems the number of trajectories required will be reduced when this method is employed.

Chapter III

Applications of the Monte Carlo Method in Neutron Physics

1. THE MONTE CARLO METHOD IN ELEMENTARY PARTICLE PROBLEMS

1. Introduction

The result of interaction between a large number of particles must be taken into account in many physical problems. The laws of elementary interactions (microscopic laws) are known from experiment or can be predicted theoretically. However, the macroscopic characteristics of matter (for example, the density) must be known for practical purposes.

The classical method of solution of such problems is based on equations which are satisfied by the macroscopic characteristics. Diffusion equations are an example of this type of approach, and well developed numerical methods are available for their solution. In other problems it is necessary to use the equations of transport, or the kinetic equations, for which numerical methods of solution have been developed only in simple cases. Finally, there are many problems for which macroscopic equations are not available at all.

The Monte Carlo method may be used for the approximate calculation of the required characteristics without recourse to macroscopic equations.

The method of macroscopic solutions will, in general, yield a considerable amount of information, for example, the asymptotic behavior, approximate relationships, and so on. The Monte Carlo method is a numerical one and should be compared with the numerical methods for the solution of macroscopic equations rather than with the method of macroscopic equations itself. In many problems, and particularly in complicated problems, the Monte Carlo method has many advantages as compared with the classical numerical methods, apart from the fact that in many cases the macroscopic equations are not known.

2. Simulation of a Physical Process

Suppose a beam of particles enters a region G of a medium. Consider one of these particles and suppose that the distribution of free paths is known, so that it is possible to choose a free path and find the point of collision of the particle with an atom in the region G.

If the region G contains a mixture of media, then it is possible to determine with which particular kind of atom the particle interacts, because the probabilities of collision are proportional to the amounts of the various atoms present.

The incident particle may cease to exist, i.e., it may become absorbed, or may be scattered, i.e., it may assume a different energy and direction of motion. If the particle is a neutron, then a collision with a fissile nucleus may give rise to a disintegration of the nucleus and to the appearance of a number of new neutrons. If the particle is a photon, it may give rise to the appearance of an electron-positron pair, or to the appearance of a free electron (photoelectric effect), and so on. The probabilities of the various interactions between incident particles of a particular kind and the target atoms are known: they are characterized by the so-called interaction cross sections, or partial cross sections. It follows that the type of interaction occurring in a given collision can also be determined.

The history of each new particle can be followed in a similar way. As a result, a branched trajectory is produced in the region G which is sometimes called a cascade or shower. Figure 13 shows this type of effect, and is, for the sake of simplicity, confined to two dimensions. It is, in a sense, the genealogical tree (p-photon, e^--electron, e^+-positron). In general, such showers will be infinite, although in practice a finite number of branches is taken. The point at which the shower must be cut off is usually clear from the conditions specifying the particular problem.

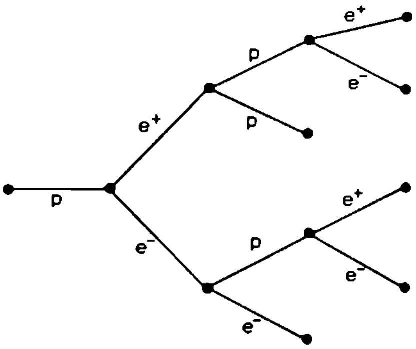


Fig. 13.

Consider now a number of incident particles large enough to give a good representation of the part of the incident beam which enters the region G, and let us construct for each of the particles a diagram of the form shown in Fig. 13. All the characteristics which are of interest may be approximately determined from this set of diagrams, for example, the number of particles of a given type at different instants of time, the number of interactions of a particular type, the energy and angular distributions of the particles, the amount of liberated energy, and so on.

This is the basic scheme for the application of the Monte Carlo method in elementary particle physics. A more detailed analysis would require the inclusion of elementary (microscopic) laws which are characteristic of the particular type of problem.

In the present chapter we shall consider some problems in neutron physics, which is the main field of application of the Monte Carlo method in physics. However, the methods which will be described in Section 3 can be used, for example, in the study of the propagation of photons, the only difference being the specific elementary interactions and the laws describing them.

3. Organization of the Calculations

One of the methods is to record all the showers and then analyze them, i.e., select the required data and carry out the necessary calculations. This involves recording the coordinates and times of all collisions, types of particles, their origin, velocities, energies, and so on. The advantage of this approach is that it will conserve all the information implicit in the data. The disadvantage of the method lies in the fact that it is necessary to store a large amount of data (problems in which the number of showers is less than 50-100 are not frequently encountered).

Since modern electronic computers have a relatively small internal (fast) memory, the above method is not as a rule convenient for them. In actual fact each shower is constructed independently of all others. The analysis is carried out for each shower in turn so that the results of the analysis gradually accumulate. In this method the number of operations involving the external (slow) memory is considerably reduced. Moreover, it is not as a rule always necessary to construct the shower first and then analyze it. It is possible to carry out the analysis as the shower grows, so that the internal memory need not store the entire shower. The following are two methods for such gradual analysis of a shower.

A. Genealogical analysis. This method is convenient when the branches are long but the degree of branching is not too high.

Particles of a given generation are used to determine the next generation and the constructed part of the shower is analyzed. This means that not more than two generations need be stored in the memory.

It should be noted that the term generation may refer to any particles which are not each other's "ancestors." For example, the particles joined by the dashed line *a* in Fig. 14 may be looked upon as belonging to a given generation. The next generation is then unambiguously determined and is indicated by *b*.

B. Lexicographic analysis of a shower. This method is convenient when the branches are not too long but the degree of branching is large. In this method a particular branch is followed right to the end and all the side branches are noted. Having reached the end of the particular branch the path is retraced to the preceding site and the final section is "cancelled." The procedure is then repeated, starting from any of the recorded sites. If, at a particular point, there are no recorded branches, then the path is retraced through one further section and the latter is cancelled. The count is completed when the entire shower is cancelled.

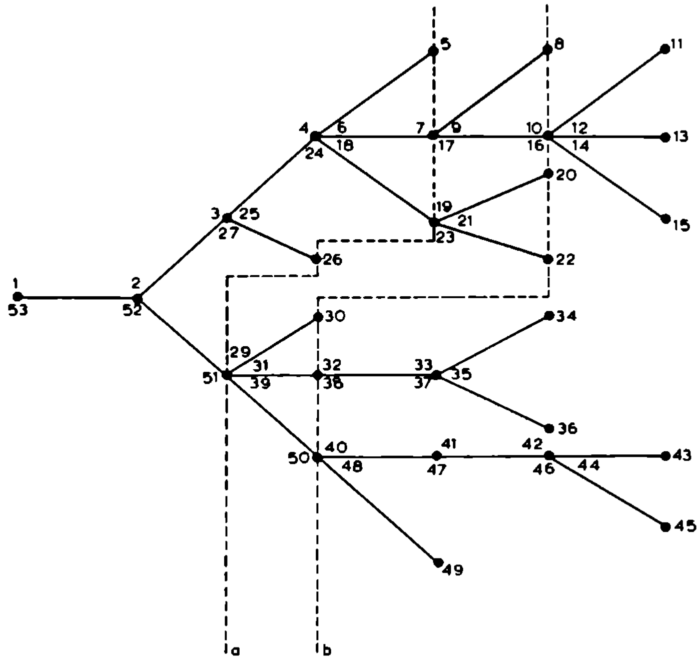


Fig. 14.

The numbers in Fig. 14 indicate the order of the lexicographic tracing of a shower. It is clear that at any given time the computer memory will store only one complete or incomplete branch of the shower, and all the offshoots along it. Figure 15 shows the part of the shower which is stored in the memory when site 11 is reached. A program which may be used with the lexicographic method has been given by Golenko [25].

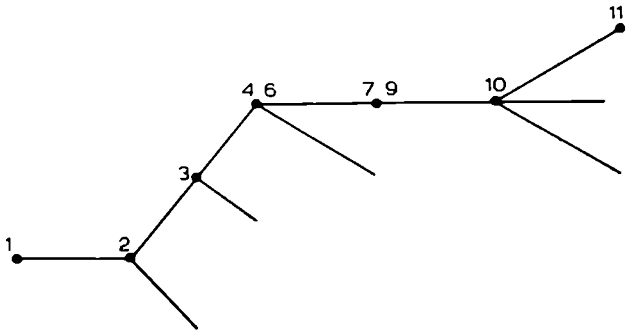


Fig. 15.

4. Estimated Accuracy

In ordinary computational practice the error in a complicated calculation can hardly ever be calculated from a theoretical formula. The accuracy must be estimated from various additional sources of information, for example, comparison of results obtained with different grids and methods, analysis of special problems for which solutions are known, and so on. Theoretical estimates of the errors (including estimates of convergence) usually play the part of additional sources of information.

A rigorous estimate of the accuracy is impossible in the case of the Monte Carlo method. However, the various additional sources of information mentioned above do remain significant. In this sense, the Monte Carlo method does not differ so very much from the classical computational methods for which a precise determination of the errors is possible in principle.

The Monte Carlo method is frequently found to have advantages even in comparison with classical methods, since even in very complicated calculations it is often possible to obtain a statistical estimate of the accuracy. In fact, most of the results obtained by the Monte Carlo method are the mathematical expectations of some random quantities ξ which are functions of the shower. The probable relative error is then given by

$$\delta_{\text{prob}}=\frac{0.67}{\mathsf{M}\xi}\sqrt{\frac{\mathsf{D}\xi}{N}},$$

where $\mathsf{D}\xi$ is the variance and N is the number of trials (showers). This quantity is a reasonable representation of the order of the statistical error (cf. Chapter II). In order to estimate $D\xi$, both $\sum_{(N)}\xi$ and $\sum_{(N)}\xi^2$ must be determined.

When the quantity under investigation is more complicated (for example, k_{eff} in Section 4) some information about the error may be obtained by examining the variation in the result with increasing N. For example, Bouquet et al. [84] assumed that the deviation from the exact value follows a normal distribution when $N \gg N_0$. However, this type of approach has not been adequately justified.

5. Introduction of Statistical Weights

Statistical weights are frequently used in the determination of particle trajectories by the Monte Carlo method (they are sometimes referred to as fictitious masses). The rigorous mathematical introduction of statistical weights (see the following paragraph) is not particularly instructive. In practice, the weights are usually introduced on the basis of elementary physical ideas which may be illustrated by an example of weights describing the escape of particles from a finite region.

Consider the trajectory of a particle in a finite convex region G. The free path l of the particle will be regarded as a random quantity with a probability density $p(x)$ which is proportional to $\exp(-x/a)$, where $0 \leqslant x < \infty$. Let x^* be the distance from the point of creation of the particle to the boundary of the region G along the direction of motion.

In the usual method of constructing a trajectory the length l is determined with the aid of the formula

$$\int_0^l p(x)\,dx = \gamma,$$

where γ represents a value of a random quantity which is uniformly distributed in the interval (0, 1). When $l > x^*$ the particle is considered to have left G. It is possible however to calculate the probability of escape of a particle from G from the formula

$$\int_{x^*}^{\infty} p(x)\,dx = e^{-x^*/a},$$

and then proceed as follows. Let the particle be replaced by a bunch consisting of a large number w of particles, so that the number of particles escaping from G is $w \exp(-x^*/a)$ while the number of particles remaining inside G is $w' = w[1 - \exp(-x^*/a)]$. The free path $\tilde{l}$ for the remainder of the packet inside G may be determined by assuming that the density remains proportional to $\exp(-x/a)$ as before, but with $0 \leqslant x \leqslant x^*$. The free path $\tilde{l}$ can then be determined from

$$\int_0^{\tilde{l}} p(x)\,dx = \gamma \int_0^{x^*} p(x)\,dx.$$

This approach is usually employed in practice. The number of particles w is in fact the statistical weight.

It is clear from the above discussion that instead of a single particle with a statistical weight w it is possible (beginning with any point on the trajectory) to take m similar particles with weights w/m and determine the history of each of them by the above method. In other words, it is always possible to branch out the trajectory and increase the number of particles.

The converse procedure is also possible. It involves the consideration of a fictitious collision in which a particle "disappears" with a probability $1 - \frac{1}{m}$ or assumes a weight mw with a probability $\frac{1}{m}$. This device is used when the statistical weight of the particle becomes so small that it cannot conveniently be followed.

Another example of the introduction of weights with the aid of the above methods is given in paragraph 14.

6. Mathematical Theory of Statistical Weights

The mathematical basis for the introduction of statistical weights is the method of selective sampling (cf. Chapter II). In fact, suppose that we are concerned with the random quantity ξ which has a density $p(x)$, $a \leqslant x \leqslant b$. It is required to determine the average value of a function $f(\xi)$ in the interval $[a, b]$:

$$\mathsf{M} f(\xi) = \int_a^b f(x)\,p(x)\,dx.$$

Instead of ξ one can take some other random quantity η which has a positive density $\tilde{p}(x)$, $a \leqslant x \leqslant b$. This will involve the calculation of the average value of another function, i.e.,

$$\tilde{f}(\eta) = f(\eta)\frac{p(\eta)}{\tilde{p}(\eta)},$$

since

$$\mathsf{M}\tilde{f}(\eta) = \int_a^b \tilde{f}(x)\,\tilde{p}(x)\,dx = \int_a^b f(x)\,p(x)\,dx = \mathsf{M} f(\xi).$$

At the same time the variances of the two quantities are different, since

$$\mathsf{M}f^2(\xi) = \int_a^b f^2(x)\, p(x)\, dx,$$

while

$$\mathsf{M}\tilde{f}^2(\eta) = \int_a^b f^2(x) \left[\frac{p(x)}{\tilde{p}(x)} \right] p(x)\, dx.$$

If $p(x)/\tilde{p}(x) < 1$ in the part of the interval (a, b) which contributes most to $\mathsf{M}f^2(\xi)$, then it is to be expected that the replacement of ξ by η will lead to a reduction in the variance and possibly to an increase in the accuracy.

In simulating the particle trajectory, it is necessary to determine the successive random quantities $\xi_1, \xi_2, \ldots$ with densities $p_1(x), p_2(x), \ldots$,* and the resulting trajectories are used to estimate the average value of a function $f(\xi_1, \xi_2, \ldots, \xi_m)$. At each stage of construction of the trajectory the determination of ξ_k may be replaced by the determination of any other random quantity η_k with positive density $\tilde{p}_k(x)$. This will yield other trajectories and the function to be averaged will have to be replaced by

$$f(\eta_1, \eta_2, \ldots, \eta_m) \frac{p_1(\eta_1)}{\tilde{p}_1(\eta_1)} \frac{p_2(\eta_2)}{\tilde{p}_2(\eta_2)} \cdots \frac{p_m(\eta_m)}{\tilde{p}_m(\eta_m)}.$$

Thus, the "weight" of the particle varies in accordance with the law

$$w_k = w_{k-1} \frac{p_k(\eta_k)}{\tilde{p}_k(\eta_k)}, \quad w_0 = 1,$$

and the averaged function $f(\eta_1, \eta_2, \ldots, \eta_m)$ must be given a "weight" w_m.

The above analysis leads to the following practical recommendation: if the weight decreases as the particle moves along the trajectory, then the method of weighting is suitable.

---

*Strictly speaking these are the conditional probabilities

$$p_k(x) = p_k(x \mid \xi_1, \ldots, \xi_{k-1}).$$

However, we shall not dwell on the more complicated concepts of the theory of probability (Markov chains).

In the example discussed in the preceding paragraph

$$p(x)=\frac{1}{a}\,e^{-x/a},\quad \tilde{p}(x)=\frac{e^{-x/a}}{a(1-e^{-x^*/a})},$$

and hence,

$$\frac{p(\tilde{l})}{\tilde{p}(\tilde{l})}=1-e^{-x^*/a}<1.$$

It should, however, be noted that in this example the intervals in which $p(x)$ and $\tilde{p}(x)$ are defined are different. The analysis given in the present section will apply, however, because the averaged functions vanish for $x>x^*$, i.e., outside the region G.

2. SIMPLE INTERACTIONS OF NEUTRONS WITH NUCLEI AND THEIR SIMULATION

7. Effective Neutron Cross Sections

In neutron physics the interaction of nuclei with neutrons is usually described by the so-called effective cross sections.

Suppose that a homogeneous beam of neutrons is incident normally on a monatomic layer of unit area, and let n be the number of atoms in the layer per unit area. If the fraction of neutrons taking part in an interaction is d then the effective cross section σ of the nucleus for the particular interaction is given by $\sigma=\frac{d}{n}$.

The effective cross sections of nuclei (sometimes referred to as the microscopic cross sections) for different interactions with neutrons are available in the literature [2]. As a rule, these cross sections depend on the neutron energy. The most frequently encountered cross sections are the following:

$\sigma_s(E)$ — scattering cross section, $\sigma_s=\sigma_{se}+\sigma_{sl}$,
$\sigma_{se}(E)$ — elastic scattering cross section,
$\sigma_{sl}(E)$ — inelastic scattering cross section,
$\sigma_c(E)$ — capture cross section,
$\sigma_f(E)$ — fission cross section,
$\sigma_a(E)$ — absorption cross section, $\sigma_a=\sigma_c+\sigma_f$,
$\sigma_t(E)$ — total cross section, $\sigma_t=\sigma_s+\sigma_c+\sigma_f$.

The macroscopic cross sections are defined as the products

$$\Sigma=\rho\sigma,$$

where ρ is the nuclear density, i.e., the number of nuclei per unit volume. σ is usually measured in barns where 1 barn $=10^{-24}$ cm^2. It follows that the dimensions of Σ are cm^{-1}.

An example of the calculation of Σ_t is given in paragraph 12. In the case of a mixture of m different media

$$\Sigma = \Sigma_{(1)} + \Sigma_{(2)} + \dots + \Sigma_{(m)}.$$

The ratios of the various cross sections to the total cross section characterize the probabilities of the various interactions which occur when a neutron collides with the nucleus. They are used in the determination of the history of the neutron. For example, when the neutron collides with the nucleus for which

$$\Sigma_t = \Sigma_s + \Sigma_c + \Sigma_f,$$

the probabilities of scattering, capture and fission are respectively equal to

$$\frac{\Sigma_s}{\Sigma_t}, \quad \frac{\Sigma_c}{\Sigma_t}, \quad \frac{\Sigma_f}{\Sigma_t},$$

i.e., the lengths of the intervals shown in Fig. 16. The order of these intervals is arbitrary, but should be fixed before the Monte Carlo method is applied. In order to determine the type of interaction, one finds the value of a random number γ and determines into which of the three intervals it falls. If

$$\gamma < \frac{\Sigma_s}{\Sigma_t},$$

then scattering occurs. Capture corresponds to

$$\frac{\Sigma_s}{\Sigma_t} < \gamma < \frac{\Sigma_s + \Sigma_c}{\Sigma_t},$$

while fission corresponds to

$$\frac{\Sigma_s + \Sigma_c}{\Sigma_t} < \gamma.$$

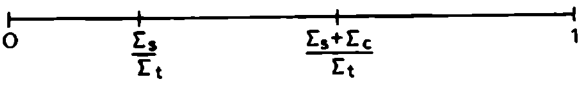

0
$\frac{\Sigma_s}{\Sigma_t}$
$\frac{\Sigma_s+\Sigma_c}{\Sigma_t}$
1

Fig. 16.

8. Elastic Scattering

Consider now the elastic scattering of a neutron by a nucleus of mass number A. This event may be defined by two random quantities which can conveniently be taken as the scattering angle $\tilde{\vartheta}$ in the system referred to the neutron-nucleus center of mass

(Fig. 17) and the azimuthal scattering angle χ, which are such that

$$0 \leqslant \tilde{\theta} \leqslant \pi, \quad 0 \leqslant \chi \leqslant 2\pi.$$

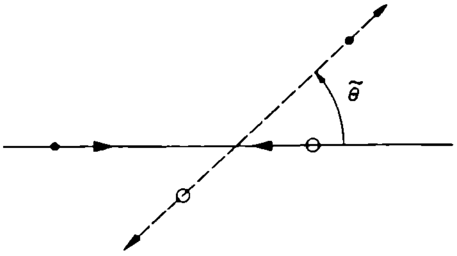


Fig. 17. Scattering of a neutron in the center of mass system.

The angle ψ through which the neutron is scattered (Figs. 18 and 19) and its final energy E' can easily be calculated from the laws of conservation of energy and momentum. They are given by

$$\cos\psi = \frac{A\cos\tilde{\theta} + 1}{\sqrt{A^2 + 2A\cos\tilde{\theta} + 1}}; \tag{3.1}$$

$$\frac{E'}{E} = \frac{A^2 + 2A\cos\tilde{\theta} + 1}{(A+1)^2}. \tag{3.2}$$

The scattering is usually assumed to be isotropic in the center of mass system. This means that $\cos\tilde{\theta}$ is uniformly distributed in the interval $(-1, +1)$ and the angle χ is distributed uniformly in the interval $(0, 2\pi)$. However, in the laboratory system of coordinates (x, y, z) illustrated in Fig. 19, the various directions of scattering are not equally probable. It is clear from (3.2) that the energy E' is uniformly distributed in the interval

$$\left(\frac{A-1}{A+1}\right)^2 E \leqslant E' \leqslant E.$$

The rule for the determination of elastic scattering which is isotropic in the center of mass system consists of the following.

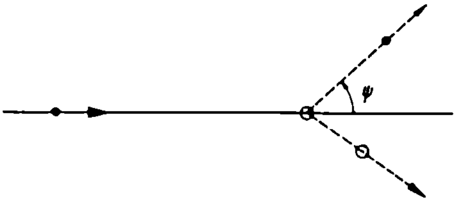


Fig. 18. Scattering of a neutron in the laboratory system.

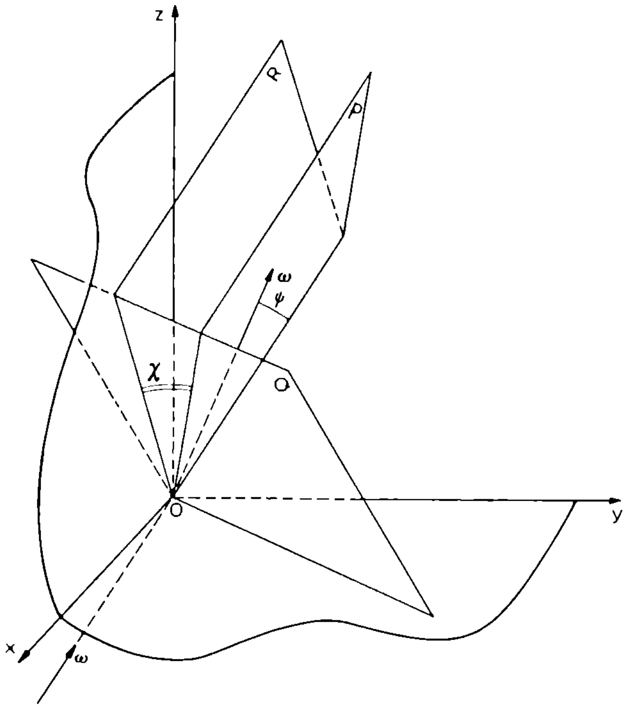


Fig. 19.

Two random quantities γ_1 and γ_2 are found and it is assumed that

$$\cos\tilde{\theta} = 2\gamma_1 - 1, \quad \chi = 2\pi\gamma_2.$$

Next, cos ψ and E' are determined from (3.1) and (3.2).*

Equations (3.1) and (3.2) remain in force in the case of anisotropic scattering. The azimuthal angle χ is again taken to be uniformly distributed in the interval $(0, 2\pi)$. However, the density of the cos $\tilde{\theta}$ distribution turns out to be proportional to the differential cross section for elastic scattering in the center of mass system $\sigma_e(\tilde{\theta})$. The formula for the determination of the quantity $\mu = \cos\tilde{\theta}$ is

$$2\pi \int_{-1}^{\mu} \sigma_e(\tilde{\theta})\, d\mu = \gamma\sigma_{se}. \tag{3.3}$$

When $\sigma_e(\tilde{\theta}) =$ const, equation (3.3) becomes identical with that given above for isotropic scattering. In fact, it follows from the normalizing condition for the probability densities that $\sigma_e(\tilde{\vartheta}) = \frac{\sigma_{se}}{4\pi}$. Substituting this into (3.3) we have $\mu = 2\gamma - 1$.

---

*Since we require not the angle χ but only cos χ, it is usual to use Neumann's assumption, which enables us to find cos χ immediately (cf. paragraph 10, formula for cos φ').

It should be noted that the differential cross section is more correctly denoted by $\sigma_e(\mu)$ since the normalization is carried out with respect to μ. The differential cross section for elastic scattering in the laboratory system $\sigma_e(\psi)$ is also used. The relation between $\sigma_e(\tilde{\vartheta})$ and $\sigma_e(\psi)$ can easily be determined with the aid of (3.1) and is

$$\sigma_e(\tilde{\vartheta})\, d\cos\tilde{\vartheta} = \sigma_e(\psi)\, d\cos\psi.$$

In elastic scattering on heavy nuclei $A \gg 1$. It follows from (3.1) and (3.2) that

$$E' \approx E, \quad \cos\psi \approx \cos\tilde{\vartheta}.$$

Experimental data show that the majority of heavy nuclei give rise to anisotropic neutron scattering, so that the corresponding differential cross sections $\sigma_e(\psi)$ are not constant. It is these cross sections which must be used in determining the directions, together with $E' = E$.

However, in many calculations the anisotropy of scattering on heavy nuclei is neglected. The scattering process is then taken to be isotropic in the laboratory system and the equations of paragraph 10 may be used to determine the direction.

9. Inelastic Scattering

The laws governing inelastic scattering are very complicated. It is usually assumed that inelastic scattering is isotropic in the laboratory system. It follows that the formulas of paragraph 10 may be used to determine the directions.

The energy distribution of inelastically scattered neutrons is usually described by empirical or semiempirical curves. The energy of an inelastically scattered neutron must also be determined probabilistically.

In the case of inelastic scattering on heavy nuclei it is often assumed that the energy distribution of the scattered neutrons is Maxwellian, i.e., the distribution density is proportional to $E' \exp(-E'/T)$ where $T = aE^{1/2}$, a is a constant which depends on the particular medium, and $0 \leqslant E' \leqslant E$. The formula for the energy of the scattered neutrons E' may be reduced to the form

$$1 - e^{-E'/T}\left(1 + \frac{E'}{T}\right) = \gamma\left[1 - e^{-E/T}\left(1 + \frac{E}{T}\right)\right],$$

and hence in each determination of the value of E' it is necessary to solve the transcendental equation

$$e^{-y}(1+y)=\varepsilon,$$

where $y=E'/T$ and $\varepsilon=1-\gamma\left[1-(1+E/T)\exp(-E/T)\right]$. When $0<\varepsilon<1$ the latter equation has a single positive root which may be computed by an iteration method:

$$y^{(0)}>0 \text{ is arbitrary;}$$

$$y^{(s+1)}=\ln(1+y^{(s)})-\ln\varepsilon.$$

10. Formulas for the Determination of the Direction of Scattering

Let $\boldsymbol{\omega}$ be a unit vector in the direction of the velocity of the neutron before the scattering event. In Cartesian coordinates

$$\boldsymbol{\omega}=\omega_1\boldsymbol{i}+\omega_2\boldsymbol{j}+\omega_3\boldsymbol{k}, \quad \omega_1^2+\omega_2^2+\omega_3^2=1,$$

while in spherical coordinates

$$\omega_1=\sin\theta\cos\varphi, \quad \omega_2=\sin\theta\sin\varphi, \quad \omega_3=\cos\theta.$$

Let $\boldsymbol{\omega}'$ be a unit vector in the direction of the velocity of the neutron after the scattering event, and let the corresponding components be ω_1', ω_2', ω_3' and the corresponding angles θ' and φ'.

The relations connecting the two sets of angles are

$$\cos\theta'=\cos\theta\cos\psi+\sin\theta\sin\psi\cos\chi,$$

$$\sin(\varphi'-\varphi)=\frac{\sin\psi\sin\chi}{\sin\theta'},$$

$$\cos(\varphi'-\varphi)=\frac{\cos\psi-\cos\theta\cos\theta'}{\sin\theta\sin\theta'}.$$

The relations between the two sets of Cartesian components are

$$\omega_3'=\omega_3\cos\psi+\sin\psi\cos\chi\sqrt{1-\omega_3^2},$$

$$\omega_2'=\frac{1}{1-\omega_3^2}\left[\omega_2(\cos\psi-\omega_3\omega_3')+\sin\psi\sin\chi\omega_1\sqrt{1-\omega_3^2}\right],$$

$$\omega_1'=\frac{1}{1-\omega_3^2}\left[\omega_1(\cos\psi-\omega_3\omega_3')-\sin\psi\sin\chi\omega_2\sqrt{1-\omega_3^2}\right].$$

In these formulas χ is the azimuthal scattering angle and ψ is the angle between the unit vectors $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$ (cf. Fig. 19 where P is the plane of scattering containing the vectors $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$, the plane Q is perpendicular to $\boldsymbol{\omega}$, the plane R contains the Oz axis and $\boldsymbol{\omega}$, and the angle χ is measured from the line of intersection of the planes Q and R).

When the scattering is isotropic in the laboratory system, the new direction ω' is independent of ω. It is convenient to determine it to begin with. This can be done as follows:

1) Take a random number γ and assume that

$$\cos\theta' = 2\gamma - 1.$$

2) Take two random numbers γ_1 and γ_2. When

$$(2\gamma_1 - 1)^2 + \gamma_2^2 \geqslant 1,$$

the particular pair selected is rejected and another is taken. When

$$(2\gamma_1 - 1)^2 + \gamma_2^2 < 1,$$

then

$$\cos\varphi' = \frac{(2\gamma_1 - 1)^2 - \gamma_2^2}{(2\gamma_1 - 1)^2 + \gamma_2^2}, \quad \sin\varphi' = \frac{2(2\gamma_1 - 1)\gamma_2}{(2\gamma_1 - 1)^2 + \gamma_2^2}.$$

In Cartesian coordinates

$$\omega_1' = \sqrt{1 - \cos^2\theta'}\cos\varphi',$$
$$\omega_2' = \sqrt{1 - \cos^2\theta'}\sin\varphi',$$
$$\omega_3' = \cos\theta'.$$

11. Fission

The number of neutrons ν which are formed during the fission process is a random quantity. As a rule only its average $\bar{\nu}$ is known. All the velocity directions are equally probable for each of the fission neutrons. The neutron energy is also a random quantity and for simplicity it is usually assumed that the spectrum of fission neutrons is independent of the energy of the incident neutron.

For example, in the case of U^{235} the average number of neutrons per fission is about 2.5, while the neutron energy distribution density is approximately given by

$$\pi(E) = \sqrt{\frac{2}{\pi e}}\sinh\sqrt{2E}\, e^{-E}$$

where the energy is in mev.

To describe the fission process by the Monte Carlo method, suppose that $m \leqslant \bar{\nu} < m + 1$ where m is an integer. In order to construct the shower, it is frequently assumed that ν takes on two values $\nu_1 = m$ and $\nu_2 = m + 1$ with probabilities $p_1 = m + 1 - \bar{\nu}$ and $p_2 = \bar{\nu} - m$. It follows that it is necessary to choose a random number γ and if $\gamma < p_1$ then it is assumed that $\nu = \nu_1$ and if $p_1 < \gamma$ then $\nu = \nu_2$. Next, for each of the fission neutrons a Monte Carlo determination is made of the directions and the energies. The formulas for equally probable directions are given in paragraph 10. The formula for the determination of the energy E' is of the form

$$\int_{E_{\min}}^{E'} \pi(E)\,dE = \gamma \int_{E_{\min}}^{E_{\max}} \pi(E)\,dE.$$

In many cases in which statistical weights are employed the random quantity ν is not determined by the Monte Carlo method. Instead, the statistical weight of the incident neutron is multiplied by $\bar{\nu}$ and the direction and energy are determined by the Monte Carlo method for this single neutron. In this way, branching trajectories can be avoided.

12. Length of the Free Path

The free path l of a neutron is a random quantity. The law of distribution of the free paths is of the form

$$\mathsf{P}\{l < x\} = 1 - e^{-\int_0^x \Sigma_t\,ds},$$

where s is the distance from the preceding collision along the direction of motion of the neutron. The distribution density is

$$p_l(x) = \Sigma_t e^{-\int_0^x \Sigma_t\,ds}.$$

In a homogeneous medium in which Σ_t is independent of s, the mean free path is given by $\bar{l} = 1/\Sigma_t$.

Example. Calculate the neutron mean free path in beryllium. It is known [2] that $\sigma_t = 6$ barn (at 300°K and energies between 0.1 and 10,000 ev). In order to determine the number of Be atoms per cubic centimeter its density ($1.84\,\mathrm{g/cm^3}$) must be multiplied by Avogadro's number (6.02×10^{23}) and the result must be divided by the mass number ($A = 9$). The final result is $\rho = 1.23 \times 10^{23}\ \mathrm{cm^{-3}}$ and hence $\Sigma_t = \rho\sigma_t = 0.74\ \mathrm{cm^{-1}}$ and $\bar{l} = 1.3$ cm.

In a homogeneous medium, the free path may be determined by the Monte Carlo method with the aid of the formula

$$l = -\frac{1}{\Sigma_t}\ln\gamma. \tag{3.4}$$

where γ is a random number. If the mean free path $\bar{l}$ is taken as the unit of length, then $l = -\ln\gamma$. If the homogeneous medium is finite and the escape of neutrons is allowed for by the appropriate statistical weight (paragraph 5) then Equation (3.4) must be replaced by

$$l = -\frac{1}{\Sigma_t}\ln\left[1-\gamma\left(1-e^{-s^*\Sigma_t}\right)\right], \tag{3.4'}$$

where s^* is the distance to the boundary of the region along the direction of flight of the neutron. The new statistical weight of the neutron is

$$w' = w\left(1-e^{-s^*\Sigma_t}\right).$$

The free path of a neutron can be defined for any type of interaction, for example, scattering. All the preceding formulas will then hold except that Σ_t must be replaced by the appropriate interaction cross section (Σ_s in the case of scattering).

13. Simulation of the Free Path in a Nonhomogeneous Medium

Consider a region consisting of a finite number of homogeneous regions. In fact, any nonhomogeneous region may be divided into small regions within which the composition of the medium remains constant.

The boundaries of the regions are, in practice, plane, spherical or cylindrical surfaces. The intersections of any ray with such surfaces may be found with the aid of the formulas of analytical geometry. The method of calculation given below may be used to simulate the free path of a neutron for any geometry.

In order to obtain an insight into the problem, the calculations are illustrated by the example shown in Fig. 20 in which a cylindrical region I (active zone of a reactor) is surrounded by a cylindrical layer II (reflector) and two cylinders III and IV (end pieces).

Suppose that a neutron at the point r_0 escapes in the direction of the unit vector ω'. The equation of the trajectory (ray) is

$$r' = r + s\omega', \quad s \geqslant 0.$$

Rules for the determination of the free path by the Monte Carlo method:

1) Find the distance (along the ray) to all the bounding surfaces for the particular geometry. In our present example this will yield the following seven numbers:

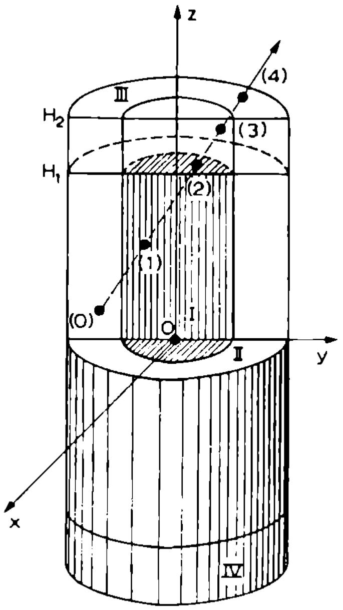


Fig. 20.

$s_1 < s_2$ — distances to the inner cylindrical surface.
s_3 — distance to the outer cylindrical surface.
s_4 — distance to the plane $z = H_1$.
s_5 — distance to the plane $z = H_2$.
s_6 — distance to the plane $z = -H_1$.
s_7 — distance to the plane $z = -H_2$.

2) Find the distance s^* (along the ray) to the outer boundary which is the smallest positive number among all the distances to the surfaces making up the outer boundary. In our example s^* is the smallest positive number among the numbers s_3, s_5 and s_7.

3) Arrange all the positive values s_m which are less than s^* in ascending order:

$$0 = s_{(0)} < s_{(1)} < s_{(2)} < \ldots < s_{(k)} = s^*,$$

and calculate the lengths of the following sections along the ray:

$$l_{(j)} = s_{(j)} - s_{(j-1)}, \qquad j = 1, 2, \ldots, k.$$

Each of these sections lies wholly inside one of the regions.

4) Find the values $\Sigma_t = \Sigma_{(j)}$ which correspond to $l_{(j)}$. To do this, it is sufficient to determine the region to which the point

$$r_{(j)} = r_0 + 0.5\,(s_{(j-1)} + s_{(j)})\,\omega'$$

corresponds.*

5) Select a value of a random number γ and find the number i for which

$$\Sigma_{(1)} l_{(1)} + \dots + \Sigma_{(i-1)} l_{(i-1)} < -\ln\gamma < \Sigma_{(1)} l_{(1)} + \dots + \Sigma_{(i)} l_{(i)}.$$

The free path is then given by

$$l = l_{(1)} + \dots + l_{(i-1)} + \frac{-\ln\gamma - (\Sigma_{(1)} l_{(1)} + \dots + \Sigma_{(i-1)} l_{(i-1)})}{\Sigma_{(i)}}.$$

If there is no value i for which this condition is satisfied, i.e., if

$$\Sigma_{(1)} l_{(1)} + \dots + \Sigma_{(k)} l_{(k)} < -\ln\gamma,$$

then the neutron is considered to have escaped from the region.

Having determined the length l, the new position of the neutron is found from

$$r_1 = r_0 + l\omega'.$$

In our example, for the ray shown in Fig. 20 $s^* = s_5$. Next,

$$s_{(1)} = s_1, \quad s_{(2)} = s_4, \quad s_{(3)} = s_2, \quad s_{(4)} = s^*.$$

The section $l_{(1)}$ belongs to II, $l_{(2)}$ belongs to I, $l_{(3)}$ belongs to III and $l_{(4)}$ also belongs to the region III.

The number of calculations involved in the above scheme may frequently be reduced by allowing for the specific properties of the particular geometry. Thus, if $s_5 > 0$ in the example considered above, then s_7 need not be calculated at all, and so on. However, this leads to a complication of the logical scheme of the calculations and therefore the programming is also more complicated (except for a very simple geometry).

Consider now the determination of the free path of a neutron when the possibility of escape is allowed for by means of statistical weights (paragraph 5). In this case the scheme will also consist of rules 1, 2, 3, and 4 and thereafter the procedure is as follows:

5) Select one of the successive values of a random quantity and calculate the quantity

$$\Gamma = 1 - \gamma\left[1 - e^{-(\Sigma_{(1)} l_{(1)} + \dots + \Sigma_{(k)} l_{(k)})}\right].$$

---

*It is clear from this that the programs should contain a part which determines the range to which any selected point r belongs.

Find the number l which is such that

$$\Sigma_{(1)}l_{(1)}+\ldots+\Sigma_{(l-1)}l_{(l-1)}<-\ln\Gamma<\Sigma_{(1)}l_{(1)}+\ldots+\Sigma_{(l)}l_{(l)}.$$

The free path is then given by

$$l=l_{(1)}+\ldots+l_{(l-1)}+\frac{-\ln\Gamma-(\Sigma_{(1)}l_{(1)}+\ldots+\Sigma_{(l-1)}l_{(l-1)})}{\Sigma_{(l)}}.$$

The new statistical weight of the neutron is given by

$$w'=w\left[1-e^{-(\Sigma_{(1)}l_{(1)}+\ldots+\Sigma_{(k)}l_{(k)})}\right].$$

14. Statistical Weights Replacing the Monte Carlo Determination of Interactions

We shall consider the two most frequently encountered methods of introducing these weights. For simplicity, we shall assume that there is only one type of scattering (elastic or inelastic).

Consider a bunch consisting of a large number w of neutrons. During a collision $w(\Sigma_c/\Sigma_t)$ neutrons will be absorbed, $w(\Sigma_s/\Sigma_t)$ neutrons will be scattered and $w(\Sigma_f/\Sigma_t)$ neutrons will give rise to fission as a result of which $\bar{\nu}w(\Sigma_f/\Sigma_t)$ fission neutrons will be produced. It may be supposed that after a collision there will be two neutrons, namely, a scattered neutron with a weight

$$w'=w\frac{\Sigma_s}{\Sigma_t}, \tag{3.5}$$

and a fission neutron with a weight

$$w'=w\frac{\bar{\nu}\Sigma_f}{\Sigma_t}. \tag{3.5$'$}$$

Consider now the neutron bunch again, but assume that during the entire free path neutrons are continuously removed from the bunch as a result of absorption. Since the probability that a neutron will be absorbed in the interval $(x,\ x+dx)$ is equal to $\Sigma_c\exp(-\Sigma_c x)\,dx$, it follows that the total number of neutrons which will be absorbed from the bunch in a free path will be $w(1-\exp(-\Sigma_c l))$.

As before, a fraction of the neutrons will be scattered while a further fraction will give rise to fissions. It may be considered that two neutrons will be produced, one of which will be the scattered neutron with a weight

$$w'=we^{-\Sigma_c l}\frac{\Sigma_s}{\Sigma_s+\Sigma_f} \tag{3.6}$$

while the other will be the fission neutron with a weight

$$w'=we^{-\Sigma_c l}\frac{\bar{\nu}\Sigma_f}{\Sigma_s+\Sigma_f}. \tag{3.6$'$}$$

In the second method of calculation, the path length is determined from $\Sigma_s + \Sigma_f$ and not from Σ_t.

In both cases considered above, the type of interaction occurring during the collision is not determined, and the trajectory branches out into two each time (if it were desired to allow for both elastic and inelastic scattering, then the trajectory would have to be branched into three each time).

When the medium does not contain fissile material ($\Sigma_f = 0$), formulas (3.5) and (3.6) become very simple. It is then unnecessary to carry out a Monte Carlo calculation of the absorption and there is no branching.

3. PASSAGE OF NEUTRONS THROUGH A PLATE

The present section will be concerned with the passage of a uniform beam of neutrons through a plane-parallel plate. It will be assumed that the plate is homogeneous and contains no fissile materials so that the total cross section is given by $\Sigma_t = \Sigma_c + \Sigma_s$. This type of problem is often encountered in practice, for example, in the design of reactor shielding, and a large number of papers have been devoted to it. In different problems it is necessary to determine the number of transmitted neutrons, their energy distribution, the reflecting power of the plate (albedo), the energy distribution of the absorbed radiation, and so on.

15. Simulation of Physical Trajectories

Let the Oz axis be perpendicular to the plane of the plate, and consider a point z such that $0 \leqslant z \leqslant h$, with the neutrons incident onto the $z = 0$ plane. The calculation is particularly simple because of the one-dimensional geometry. The state of a neutron is characterized by only three quantities, namely, the z coordinate, the energy E and the direction of flight defined by $\mu = \cos\theta$ (Fig. 21). The initial values for the trajectories are selected as follows: The initial value of z is always taken to be zero, while the initial direction and energy depend on the properties of the incident beam.

Consider some special cases.

a) Monoenergetic beam; here the initial energy E_0 is given.

b) The energy spectrum $n(E)$ of the incident beam is given. The initial energy E_0 is then determined by the Monte Carlo method, using the formula

$$\int_{E_{\min}}^{E_0} n(E)\,dE = \gamma \int_{E_{\min}}^{E_{\max}} n(E)\,dE.$$

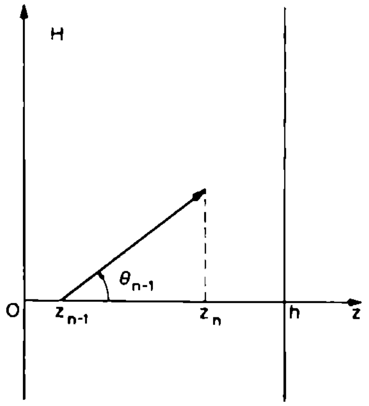


Fig. 21.

c) A directed beam; here the value μ_0 is given.

d) Spatially isotropic beam. The value of μ_0 is determined statistically from the formula $\mu_0 = \gamma$.

Subsequent calculation of the trajectory is carried out as follows.

1) The coordinate of the nth collision ($n = 1, 2, \ldots$) is determined statistically from

$$z_n = z_{n-1} - \frac{\mu_{n-1} \ln \gamma}{\Sigma_t (E_{n-1})}.$$

2) The following conditions are then verified: (a) If $z_n - h > 0$, the neutron has passed through the plate, and (b) if $z_n < 0$, the neutron has been back-scattered. The trajectory terminates (and a count is recorded in the corresponding counter) when either of these two conditions is satisfied.

3) If $0 \leqslant z_n \leqslant h$ then the fate of the neutron after the collision is determined statistically by taking successive values of a random quantity γ. If $\gamma < \frac{\Sigma_c}{\Sigma_t}$ then the neutron is considered to have been absorbed and its trajectory terminates (this is recorded in the corresponding counter).

4) If $\frac{\Sigma_c}{\Sigma_t} < \gamma$ then the neutron undergoes a scattering event and its direction and energy must be determined statistically.

In particular, if the scattering is elastic and isotropic in the center of mass system then

$$\cos \tilde{\theta} = 2\gamma - 1, \quad \chi = 2\pi\gamma_1,$$

$$\cos \varphi = \frac{1 + A \cos \tilde{\theta}}{\sqrt{1 + 2A \cos \tilde{\theta} + A^2}},$$

$$E_n = E_{n-1} \frac{A^2 + 2A \cos \tilde{\theta} + 1}{(A+1)^2},$$

$$\mu_n = \mu_{n-1} \cos \psi + \cos \chi \sqrt{(1 - \mu_{n-1}^2)(1 - \cos^2 \psi)}.$$

Thus, a neutron in the state $(z_{n-1}, \mu_{n-1}, E_{n-1})$ undergoes a transition to the state (z_n, μ_n, E_n). The calculation of the trajectory is continued until one of the three conditions for its termination (transmission, reflection or absorption) is satisfied. Next, the values of z_0, μ_0, E_0 are again selected (or are statistically determined) and are used to calculate the next trajectory.

16. The Use of Statistical Weights Representing Absorption

When the fraction of transmitted neutrons is small (this is a typical case in shielding design) then it is clear that the above method will not be very efficient. For example, if the probability of transmission is of the order of 10^{-5} then in order to obtain, say, 10 transmitted neutrons, it is necessary to construct about 10^6 trajectories, and the probable error in the determination of the probability of transmission will be about 20%. In such cases the simulation of physical trajectories is inconvenient and statistical weights should be introduced instead.

Each neutron state is then described by four quantities (z_n, μ_n, E_n, w_n) and the initial values z_0, μ_0, E_0 are chosen as in paragraph 15 with w_0 usually assumed to be equal to unity. Subsequent calculation of the trajectory is continued as in paragraph 15, except for step (3); the absorption of the neutron is not determined statistically and instead its weight is adjusted so that

$$w_n = w_{n-1} \frac{\Sigma_s(E_{n-1})}{\Sigma_t(E_{n-1})}.$$

This method of computation allows us to conserve a large number of trajectories when the absorption is large.

17. Calculation of the Probability of Transmission

Suppose that it is desired to determine the probability $p(h)$ of the transmission of neutrons through a parallel plate. In the case of physical simulation (paragraph 15) a count is assumed to occur in the counter whenever a trajectory ends in transmission $(z_n - h > 0)$. When, out of a total number N of such "case histories," the number resulting in transmission is N' then the required probability is given by

$$p(h) \approx \frac{N'}{N}. \tag{3.7}$$

In the calculation involving the statistical weights (paragraph 16) the counter reading is increased by w_{n_j} which are the weights of the transmitted neutrons. The required probability is then given by

$$p(h) \approx \frac{\sum\limits_{j=1}^{N}{}^{*} w_{n_j}}{\sum\limits_{j=1}^{N} w_0}, \tag{3.8}$$

where $\sum^*$ is evaluated over all the transmitted neutrons and $\sum$ over all the trajectories; $j = 1, 2, \ldots, N$ are the serial numbers of the trajectories, and $n = 0, 1, 2, \ldots, n_j$ are the serial numbers of the collisions inside the plate for the trajectory with number j.

When $w_0 = 1$, the denominator in (3.8) is equal to N.

In order to illustrate why the second method is as a rule better than the first, let us consider two simple random quantities ξ_n and η_n. Let Q_n be the probability of transmission without collisions for a neutron from the nth state, which is assumed to be known. This probability can easily be calculated from the formulas given in paragraph 12:

$$Q_n = \begin{cases} e^{-\Sigma_t(E_n)\frac{h-z_n}{\mu_n}}, & \text{where } \mu_n > 0; \\ 0, & \text{where } \mu_n \leqslant 0. \end{cases}$$

The random quantity corresponding to the first method of calculating $p(h)$ is

$$\xi_n = \begin{cases} 1 & \text{with the probability } (\Sigma_s/\Sigma_t)\,Q_n, \\ 0 & \text{with the probability } 1 - (\Sigma_s/\Sigma_t)\,Q_n, \end{cases}$$

while the random quantity corresponding to the second method is

$$\eta_n = \begin{cases} \frac{\Sigma_s}{\Sigma_t} & \text{with the probability } Q_n, \\ 0 & \text{with the probability } 1 - Q_n. \end{cases}$$

The mathematical expectations of these quantities are

$$\mathrm{M}\xi_n = \mathrm{M}\eta_n = \left(\frac{\Sigma_s}{\Sigma_t}\right) Q_n.$$

In order to compare the variances, it is sufficient to compare the mathematical expectations of the squares

$$\mathrm{M}\xi_n^2 = \left(\frac{\Sigma_s}{\Sigma_t}\right) Q_n,$$

$$\mathrm{M}\eta_n^2 = \left(\frac{\Sigma_s}{\Sigma_t}\right)^2 Q_n.$$

It is clear that $\mathsf{M}\eta_n^2 \leqslant \mathsf{M}\xi_n^2$ and hence*

$$\mathsf{D}\eta_n \leqslant \mathsf{D}\xi_n.$$

It can be seen from the above discussion that it would be still more convenient to carry out a direct calculation of the quantity $(\Sigma_s/\Sigma_t)\,Q_n$ for each collision separately. We are thus led to a third method of estimating $p(h)$:

$$p(h) \approx \frac{\sum\limits_{j=1}^{N} \sum\limits_{n=0}^{n_j} w_n Q_n}{\sum\limits_{j=1}^{N} w_0}. \tag{3.9}$$

The latter formula involves a considerable volume of calculation for each collision. Nevertheless it does lead to a considerable economy as compared with (3.8) if the same accuracy of the final result is required. By using (3.9), more information is extracted from each collision.

18. Further Remarks on the Use of Trajectories

In constructing a trajectory with a statistical weight it is not essential to continue the trajectory until the neutron leaves the layer. The trajectory may be terminated when the weight w_n becomes smaller than some given number ε which is, of course, chosen on the basis of physical considerations. Other conditions for terminating the trajectories may arise in various specific problems. For example, if we are interested in the absorption of neutrons within a given spectral band, then the trajectory may be terminated as soon as the neutron energy leaves the energy band.

During the calculation of the trajectory, a considerable amount of useful information may be extracted from each collision. For example, in order to find the probability $q(h)$ of reflection of neutrons by the plane-parallel plate, one can compute the reflection probability

$$\overline{Q}_n = \begin{cases} e^{\Sigma_t(E_n)\frac{z_n}{\mu_n}}, & \text{where } \mu_n < 0, \\ 0, & \text{where } \mu_n \geqslant 0. \end{cases}$$

---

*The quantities ξ_n (and also η_n) corresponding to different n are not independent. It follows that this demonstration of the advantages of this second method as compared with the first method is not rigorous.

The formula for $q(h)$ analogous to (3.9) is

$$q(h) \approx \frac{\sum_{j=1}^{N} \sum_{n=0}^{n_j} w_n \overline{Q}_n}{\sum_{j=1}^{N} w_0}.$$

At the same time, it is possible to calculate the probability of transmission (or reflection) for a plate of some other smaller thickness $h' < h$. For the thinner plate the neutron may be regarded as transmitted when $z_n - h' > 0$. Even when subsequent collisions result in the return of the neutron into the thinner plate it need not be taken into account. For example, the neutron whose trajectory is shown in Fig. 22 is reflected by the thicker plate after $n_j = 8$ while for the thinner plate it is regarded as transmitted after $n'_j = 3$.

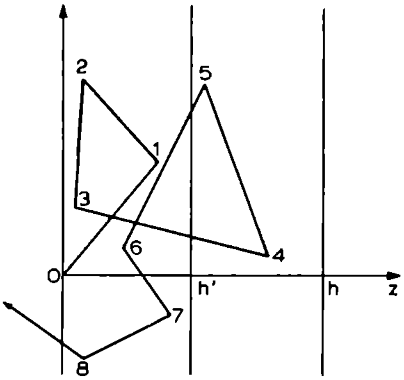


Fig. 22.

It follows that

$$p(h') \approx \frac{\sum_{j=1}^{N}{}^{*\prime}\, w_{n'_j}}{\sum_{j=1}^{N} w_0},$$

where $\sum^{*\prime}$ is taken over all the escapes from the thinner plate and n'_j is the serial number of the last collision in this plate. It is, of course, possible to use a formula of the form of (3.9):

$$p(h') \approx \frac{\sum_{j=1}^{N} \sum_{n=0}^{n'_j} w_n Q'_n}{\sum_{j=1}^{N} w_0},$$

where

$$Q'_n = \begin{cases} e^{-\Sigma_t(E_n)\frac{h'-z_n}{\mu_n}} & \text{for } \mu_n > 0, \\ 0 & \text{for } \mu_n \leqslant 0. \end{cases}$$

All these formulas are, in fact, special cases of (3.10). It is true that in paragraph 20 the trajectories are constructed using Σ_s and not Σ_t and absorption is allowed for by weights of a different type (cf. paragraph 14), but this does not alter the basic situation.

19. The Method of Similar Trajectories

Morton [210] has suggested another method whereby trajectories constructed for a homogeneous plate H may be used in calculations for a similar plate H' of thickness $h' = \alpha h$. He has used this method to obtain the curve $p = p(h)$ for a particular problem, and has estimated the derivative $\left(\frac{dp}{dh}\right)_{h=0}$.

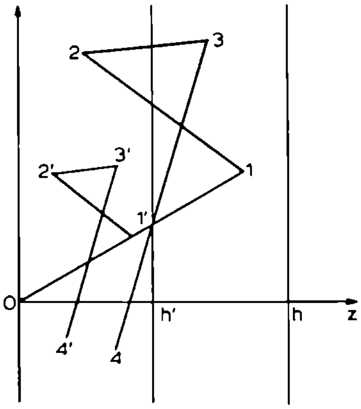


Fig. 23.

By altering the scale by a factor of α, i.e., by assuming $z' = \alpha z$, each trajectory in H may be put into one-to-one correspondence with a trajectory in H' (Fig. 23). The directions and energies for the new trajectory are already known statistically and may be associated with the corresponding collisions. However, the lengths of the sections of the broken curve were determined statistically on the old scale. According to paragraph 6 this distortion may be compensated for by the introduction of additional weights. This, in fact, is the principle of the method of similar trajectories.

In addition to the new weights w'_n it is convenient to introduce the weights v'_n with which the neutron reaches the point z'_n. The weights w'_n correspond to the weights w_n which were used to determine the original trajectories in H and which allow for the absorption. It may be considered that in H, $v_n = w_{n-1}$ but $v'_n \neq w'_{n-1}$, since v'_n takes into account the distorted statistical determination of the path length.

Thus, consider the first section of the broken curve in H which has a length

$$z_1 - z_0 = \mu_0 l_0,$$

where l_0 is chosen from the density

$$\varphi(x) = \Sigma_0 e^{-\Sigma_0 x}$$

(for the sake of brevity $\Sigma_t(E_0)$ will be replaced by Σ_0). For the broken curve in H' the length of the first section is

$$z'_1 - z'_0 = \mu_0 l'_0,$$

where l'_0 is also chosen from the density

$$\varphi(x) = \Sigma_0 e^{-\Sigma_0 x}.$$

However, instead of this statistical calculation it was assumed that $l'_0 = \alpha l_0$. Since the density of the product αl_0 is

$$\varphi_1(x) = \left(\frac{\Sigma_0}{\alpha}\right) e^{-\left(\frac{\Sigma_0}{\alpha}\right)x},$$

it follows that the weight which must be introduced (paragraph 6) is equal to

$$\frac{\varphi(l'_0)}{\varphi_1(l'_0)} = \alpha e^{(1-\alpha)\Sigma_0 l_0}.$$

Assuming that $w'_0 = w_0$ we have

$$v'_1 = w_0 \alpha e^{(1-\alpha)\Sigma_0 l_0},$$
$$w'_1 = w_1 \alpha e^{(1-\alpha)\Sigma_0 l_0}.$$

The formulas for the new weights may be derived in a similar way:

$$v'_n = f_n w_{n-1}, \quad w'_n = f_n w_n,$$

where

$$f_n = \alpha^n e^{(1-\alpha)[\Sigma_t(E_0) l_0 + \dots + \Sigma_t(E_{n-1}) l_{n-1}]}.$$

In order to determine f_n one can use $f_0 = 1$ and the recurrence formula

$$f_n = f_{n-1} \alpha e^{(1-\alpha)\Sigma_t(E_{n-1}) l_{n-1}}.$$

Formulas for estimating the probability of transmission of neutrons $p(h)$ through the plate H are given in paragraph 17. In order to establish the analogous formulas for H' we note that when $\mu_n > 0$ we have

$$Q'_n = e^{-\Sigma_t(E_n)\frac{h'-z'_n}{\mu_n}} = e^{-\alpha\Sigma_t(E_n)\frac{h-z_n}{\mu_n}} = (Q_n)^\alpha.$$

Bearing this in mind we have*

$$p(h') \approx \frac{\sum\limits_{j=1}^{N}{}^{*} f_{n_j+1} w_{n_j}}{\sum\limits_{j=1}^{N} w_0}$$

and correspondingly

$$p(h') \approx \frac{\sum\limits_{j=1}^{N} \sum\limits_{n=0}^{n_j} f_n w_n Q_n^\alpha}{\sum\limits_{j=1}^{N} w_0}.$$

The method of similar trajectories may be used to carry out simultaneous calculations for a series of plates: after each collision a series of weights is determined (for all the required α) and these are accumulated as the trajectories are completed.

Suppose now that it is necessary to calculate the transmission of neutrons through plates having thicknesses $h_1, h_2, \ldots, h_k$. The problem arises as to which is the basic plate. There are grounds for assuming that the most convenient procedure is to take the plate with the greatest thickness, so that all the factors α are less than unity. The accuracy of the calculations will then be of the same order of magnitude for all the plates.

20. Simulation of Collision Density

It is usually assumed that a large internal memory is not required in computers used for Monte Carlo calculations since each trajectory is simulated independently of all others. However, if a

---

*The sum of the weights of the transmitted neutrons is equal to

$$\sum\limits_{j=1}^{N}{}^{*} v'_{n_j+1} = \sum\limits_{j=1}^{N}{}^{*} f_{n_j+1} w_{n_j}.$$

large memory is available, then the collision density can be effectively simulated, i.e., all the states of all the neutrons may be stored. These can then be used to solve analytically a large number of similar problems using the same trajectories. This represents the most complete use of all the information obtained during the realization of the neutron trajectories. This type of calculation was first used by Berger et al. [78, 79] in the study of the transmission of gamma-radiation.

Suppose that a monoenergetic beam of neutrons with initial energies E_0 and angle of incidence $\theta_0 = 0$ is incident on the plane $z = 0$ from the region $z < 0$. Thus, the initial state of each neutron is $z_0 = 0$, $\mu_0 = 1$, E_0. The trajectory of each of the N neutrons will be constructed without allowance for absorption* or for the boundary conditions:

1) The coordinate of the nth scattering ($n = 1, 2, \dots$) is determined statistically:

$$z_n = z_{n-1} - \frac{\mu_{n-1}}{\Sigma_s(E_{n-1})} \ln \gamma;$$

2) The new direction μ_n and the new energy E_n are determined statistically as in paragraph 15;

3) The condition for the termination of the trajectory is verified (it is convenient to take this condition in the form $E_n < \varepsilon$, where ε is a given number).

All the states are recorded so that for the jth neutron we have the following series of states:

$$\begin{gathered} z_0(j),\ \mu_0(j),\ E_0(j); \\ z_1(j),\ \mu_1(j),\ E_1(j); \\ \cdots\cdots\cdots\cdots \\ z_{n_j}(j),\ \mu_{n_j}(j),\ E_{n_j}(j), \end{gathered}$$

where $j = 1, 2, \dots, N$. These states may be regarded as an empirical sampling from the collision density $g(z, \mu, E)$ which represents the probability density for the scattering of a neutron at the point z in the direction μ with an energy E.

Since most of the quantities which are of interest in such problems are expressed in terms of integrals containing the collision density $g(z, \mu, E)$, it follows that the approximate determination of these quantities can be reduced to the evaluation of certain sums over the recorded states.

---

*If all the states are recorded the weights can be introduced later.

Consider the quantity $p_{lk}(h)$ which is the probability of transmission of neutrons through a plate $0 \leqslant z \leqslant h$ where the direction and energy of the neutrons escaping from the plate lie in certain fixed intervals $\Delta\mu_l$, ΔE_k. A knowledge of all the $p_{lk}(h)$ is equivalent to a knowledge of the spectrum of the transmitted neutrons. Consider the sum

$$p_{lk}(h) \approx \frac{1}{N} \sum_{j=1}^{N} \sum_{n=0}^{n_j} B_n(j)\, w_n(j)\, Q_n(j,\ h)\, f_n^{l,\, k}(j). \tag{3.10}$$

In this sum: $B_n(j)$ represents the geometrical boundary conditions of the problem. In our case, $B_n(j) = 1$ when $0 \leqslant z_m(j) \leqslant h$ for all $m = 0,\ 1,\ 2,\ \ldots,\ n$; in all other cases $B_n(j) = 0$. The weights $w_n(j)$ represent absorption:

$$w_n(j) = \prod_{m=0}^{n-1} e^{-\Sigma_c(E_m(j))\, \frac{z_{m+1}(j) - z_m(j)}{\mu_m(j)}}$$

The product is equal to the probability that there will be no absorption along the entire path from the beginning of the trajectory to the nth scattering.

The function $Q_n(j,\ h)$ is the probability that a neutron in the nth state will intersect the plane $z = h$ without undergoing any further scattering. It is clear that

$$Q_n(j,\ h) = \begin{cases} e^{-\Sigma_s(E_n(j))\, \frac{h - z_n(j)}{\mu_n(j)}}, & \text{when } \dfrac{h - z_n(j)}{\mu_n(j)} > 0, \\ 0, & \text{when } \dfrac{h - z_n(j)}{\mu_n(j)} \leqslant 0. \end{cases}$$

Finally, the quantity $f_n^{l,\, k}(j)$ gives the spectral classification:

$$f_n^{l,\, k}(j) = \begin{cases} 1, \text{when } \mu_n(j) \in \Delta\mu_l,\ E_n(j) \in \Delta E_k, \\ 0 \text{ in the opposite case.} \end{cases}$$

The spectrum of neutrons reflected by the plate $0 \leqslant z \leqslant h$ is calculated in a similar way:

$$q_{lk}(h) \approx \frac{1}{N} \sum_{j=1}^{N} \sum_{n=0}^{n_j} B_n(j)\, w_n(j)\, Q_n(j,\ 0)\, f_n^{l,\, k}(j).$$

All the quantities in this formula have the same meaning as before.

Similar formulas may be written down for the neutron flux and a number of other physical quantities.

The terms corresponding to $n = 0$ in Equation (3.10) represent unscattered neutrons and can occasionally be conveniently replaced by an analytical expression.

Instead of simulating the passage of neutrons at different angles of incidence θ_0, it is possible to confine the analysis to the case $\theta_0 = 0$ and yet consider trajectories at all possible orientations. If the complete states $x_n(j)$, $y_n(j)$, $z_n(j)$, $\mu_n(j)$, $\varphi_n(j)$, $E_n(j)$ have been recorded, then it is easy to compute the transmission through a plate of given thickness h by orienting it in space at the required angle (Fig. 24).

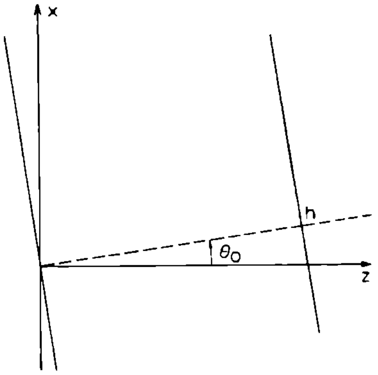


Fig. 24.

Instead of z_n and μ_n in Equation (3.10), it is necessary to use the quantities

$$\bar{z}_n = x_n \sin\theta_0 + z_n \cos\theta_0,$$

$$\bar{\mu}_n = \cos\theta_n \cos\theta_0 + \sin\theta_n \sin\theta_0 \cos\varphi_n.$$

21. A Numerical Example

In order to compare some numerical methods of solution of the transport equation, a calculation was made of the transmission of neutrons through a beryllium plate ($A = 9; \sigma_a = 0; \sigma_t = \sigma_s = \sigma_{se}$). The scattering was assumed to be isotropic in the center of mass system.* The thickness of the plate was $h = 5$ mean free paths (cf. the example in paragraph 12). The energy E was replaced by the so-called lethargy, which is defined by $u = \ln(E_0/E)$. The incident neutron flux density was assumed to be constant for $0 < \mu < 1$ and $0 < u < 12$.

The method of paragraph 20 was used to calculate the number I_{lk} of transmitted neutrons with directions and lethargies corresponding to the above intervals, for a unit area perpendicular to the direction of exit. Since the density of the computed quantity is given by

---

*The calculations were carried out by O. B. Moskalev, I. G. Krutikova, and V. A. Chuyanov under the general direction of E. S. Kuznetsov.

$$I(h, \mu, u) = \int_0^h g(z, \mu, u)\, e^{-\frac{h-z}{\mu}} \frac{dz}{\mu} + I(0, \mu, u)\, e^{-\frac{h}{\mu}},$$

the computational formula was taken in the form

$$I_{lk} \approx \frac{1}{N} \sum_{j=1}^{N} \sum_{n=1}^{n_j} \frac{1}{\mu_n} e^{-\frac{h-z_n}{\mu_n}} f_n^{l,\,k} + e^{-\frac{h}{\bar{\mu}_l}} (\mu_{l+1} - \mu_l),$$

where $\bar{\mu}_l = 0.5(\mu_l + \mu_{l+1})$. The following five intervals (groups) of values of the lethargy were considered:

$$0 \leqslant u < 3;\ 3 \leqslant u < 6;\ 6 \leqslant u < 9;\ 9 \leqslant u < 12;\ 12 \leqslant u < \infty$$

together with the following twenty intervals of values of μ:

$$\mu_{l+1} - \mu_l = 0.05.$$

Formulas for the calculation of the trajectories.* The initial values were: $z_0 = 0$, $\mu_0 = \gamma$, $u_0 = 12\gamma$. The transition from the nth collision to the $(n+1)$th collision was taken to be

$$z_{n+1} = z_n - \mu_n \ln \gamma;$$

$$\cos\chi = \frac{\gamma_1^2 - \gamma_2^2}{\gamma_1^2 + \gamma_2^2}, \text{ provided that } \gamma_1^2 + \gamma_2^2 < 1;$$

$$\cos\tilde{\theta} = 2\gamma - 1;$$

$$\cos\psi = \frac{1 + A\cos\tilde{\theta}}{\sqrt{1 + 2A\cdot\cos\tilde{\theta} + A^2}};$$

$$\mu_{n+1} = \mu_n \cos\psi + \cos\chi \sqrt{(1 - \mu_n^2)(1 - \cos^2\psi)};$$

$$u_{n+1} = u_n - \ln(1 + 2A\cos\tilde{\theta} + A^2) + 2\ln(A + 1).$$

Results (obtained with 9,000 trajectories). Figure 25 shows histograms of the angular distribution of transmitted neutrons for each of the energy groups, and also the values obtained by the multigroup method of solution of the transport equation [44].

Table 10 shows the values of the flux average with respect to the solid angle.**

---

*Use was made of the pseudo-random numbers $\{\gamma\}$ given by Sobol' [57].

**The first group $0 \leqslant u < 3$ is "too large" for the multigroup method. This explains the discrepancy between the two results in the first group.

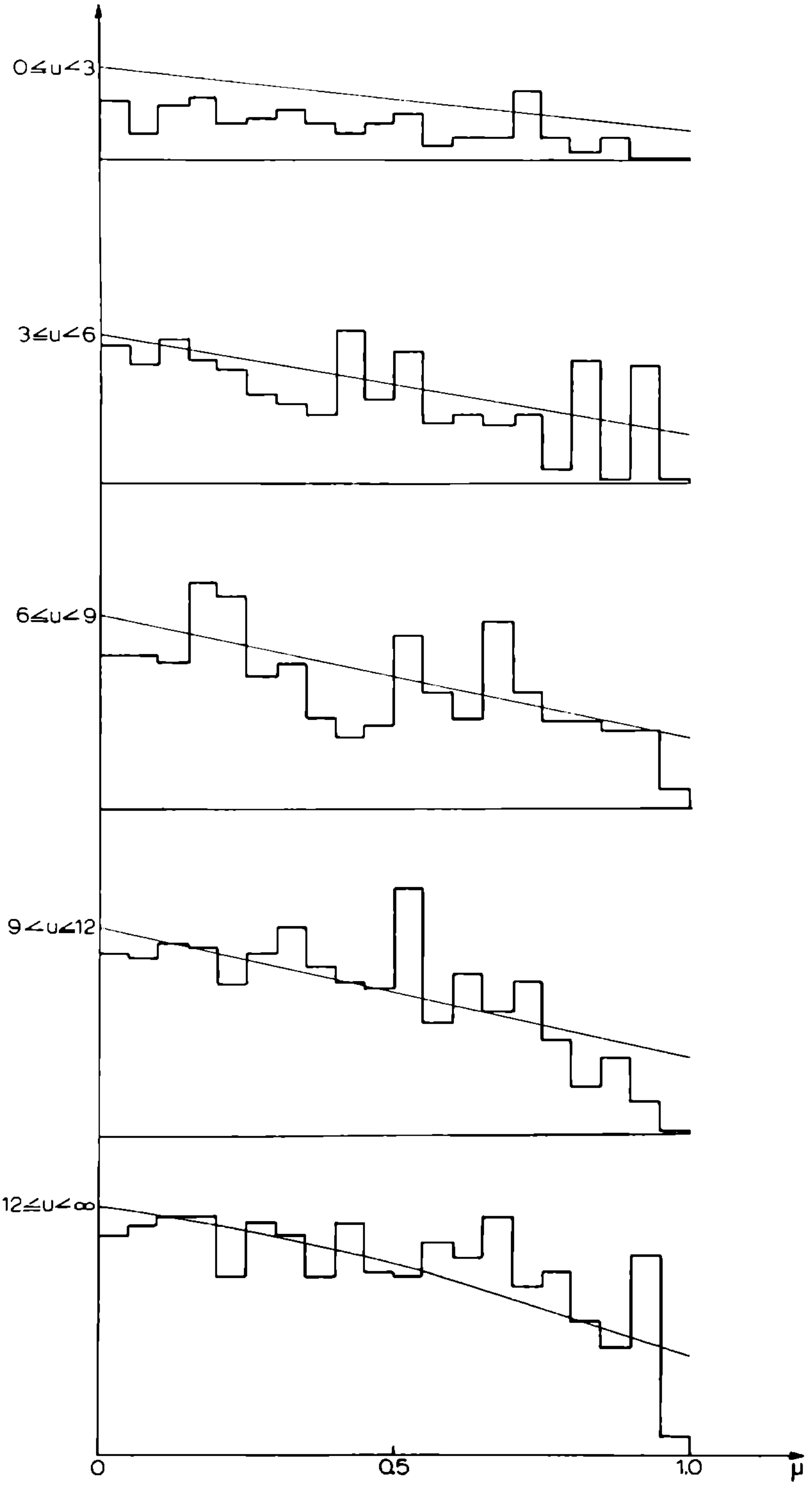


Fig. 25.

Table 10

Method	Group				
	$0 \leqslant u < 3$	$3 \leqslant u < 6$	$6 \leqslant u < 9$	$9 \leqslant u < 12$	$12 \leqslant u < \infty$
Monte Carlo. . .	0.0494	0.189	0.230	0.250	0.337
Multigroup . . .	0.1084	0.192	0.240	0.264	0.340

4. SOME METHODS OF CALCULATION OF THE CRITICALITY OF NUCLEAR REACTORS

22. Formulation of the Problem

Consider a volume R containing a fissile medium (for example, uranium). A neutron entering this region may escape from R without colliding with the nuclei, or it may interact with a nucleus and thereby become absorbed or scattered, or it may give rise to fission yielding ν new neutrons. A large number of such interactions will lead to a change in the average number of neutrons in R.

The volume R is called critical if the average number of neutrons within R remains constant. If the number of neutrons decreases, then the volume is called subcritical, while if the number of neutrons increases, the region is referred to as supercritical.

In a subcritical region a chain reaction will not be maintained. In a supercritical region the chain reaction will develop and lead to an explosion. Roughly speaking, the control of a reactor consists in the maintenance of the critical state.

Usually, the active zone of a reactor, which contains the fissile material, is surrounded by a reflector which is in the form of a medium with low neutron absorption cross section (graphite, beryllium). The presence of the reflector reduces the leakage of neutrons from the system, and ensures that the critical state is reached with a smaller amount of fuel (cf. Fig. 20). Moreover, a reactor incorporates control rods which are made of a material having a high neutron absorption cross section (boron). The control of the reactor is carried out by introducing or removing such rods from the active zone.

The effective neutron multiplication coefficient k_{eff} of a reactor is defined as the average number of secondary fission neutrons per primary fission-inducing neutron. More precisely, let n_1 be the number of all the neutrons belonging to a given generation and consider the fate of each of these neutrons prior to its escape from the reactor, or until the time when it becomes absorbed (with or without fission). If the number of second-generation neutrons associated with the original n_1 neutrons is n_2, then $k_{\text{eff}} = n_2/n_1$ (a more precise mathematical definition of k_{eff} is given in paragraph 24).

It is clear that the reactor will be in a critical state when $k_{\text{eff}} = 1$. When $k_{\text{eff}} > 1$ the state is supercritical and when $k_{\text{eff}} < 1$ it is subcritical.

23. Simulation of Physical Trajectories

Suppose that a finite number $\bar{n}_0$ of neutrons is placed in the reactor. It is desirable that the distribution should be close to the true distribution although, in fact, it can be arbitrary. Let us select a time interval Δt which is several times larger than the mean lifetime T of a neutron in the reactor. A Monte Carlo determination of the history of each of these neutrons during the time Δt will yield the number of new neutrons $\bar{n}_1$. After a sufficiently large number j of such time intervals the ratio $\bar{n}_{j+1} : \bar{n}_j$ will follow the formula

$$\bar{n}_{j+1} : \bar{n}_j \approx e^{(k_{\text{eff}} - 1)\frac{\Delta t}{T}}.$$

and its magnitude will be an indication of the criticality of the reactor.

However, the actual realization of this type of calculation turns out to be quite complicated. The true distribution of neutrons in the reactor is in general a function of six variables, namely, the components of the position vector r and the components of the velocity $\boldsymbol{v}$ (instead of the latter the unit vector $\omega = \boldsymbol{v}/v$ and the energy E of the neutron are often used).

A large number of neutrons is necessary in order to reflect the properties of this distribution correctly.

24. Determination of the Number of Generations

Suppose that the spectrum of fission neutrons is independent of the velocity of the incident fission-inducing neutron. In this case it is convenient to determine the criticality by determining the number of generations.

Let n_1 neutrons be placed in the reactor and suppose that these are the first-generation neutrons. A Monte Carlo determination of the history of each of these neutrons right up to its disappearance (capture, absorption with fission, or escape) will yield the number of second-generation neutrons. After a sufficiently large number l of generations the ratio $n_{l+1} : n_l$ will be given by

$$n_{l+1} : n_l \approx k_{\text{eff}}.$$

In this method of calculation the simulation procedure is concerned only with the distribution of fission neutrons rather than with the distribution of all the neutrons in the reactor. By definition, the true fission-neutron density (birth rate) is equal to the product $\pi(v)\ n(r)$. The spectrum of fission neutrons $\pi(v)$ is known, so that it is sufficient to simulate the density $n(r)$ which is, in general, a function of three variables. It is clear that in the reflector $n(r)=0$.

Consider now the mathematical significance of this method of calculation. Let $g(r, r')$ be the density of the average number of secondary neutrons at the point r which are produced by a single primary neutron from the point r'. The density of secondary neutrons is then given by

$$n_2(r)=\int g(r, r')\,n_1(r')\,dr'. \tag{3.11}$$

The transition from generation to generation is equivalent to an iteration of the initial density $n_1(r)$ by means of the integral operator (3.11). The kernel of this operation $g(r, r')$ is very complicated, and it is not useful to write it out analytically; it is approximately realized in the computation of the histories. The kernel is positive for the active zone.

It is now possible to give a rigorous definition of the effective neutron multiplication coefficient in a reactor. The requirement $n_2(r)=k_{\text{eff}}\,n_1(r)$ is equivalent to

$$\int g(r, r')\,n_1(r')\,dr'=k_{\text{eff}}\ n_1(r).$$

The latter equation shows that the density $n_1(r)$ should be an eigenfunction of the integral operator (3.11) and k_{eff} its eigenvalue. Since $n_1(r)\geqslant 0$ (in view of its physical significance), it follows that k_{eff} is the largest eigenvalue of the operator (3.11).

It is known from the theory of integral equations that, subject to very broad restrictions imposed on the positive kernel, the iteration of $\varphi_i(r)$ by any nonnegative function $\varphi_0(r)$ leads (with suitable normalization) to the first eigenfunction $n_1(r)$ of the kernel, and the ratios $\varphi_{i+1}(r)/\varphi_i(r)$ at each point tend to the first eigenvalue. This is the basis of the method of calculation given above.

Instead of the estimator $k_{\text{eff}}\approx n_{i+1}/n_i$, Hammersley and Morton [148] have suggested the use of a maximum likelihood method. Thus, if it is considered that when $i=s$ the ratio has already "settled down," while the count of the generations is continued up to $i=l$, then

$$k_{\text{eff}}\approx\frac{\hat{n}-n_s}{\hat{n}-n_l},$$

where $\hat{n} = n_s + n_{s+1} + \ldots + n_l$. The corresponding estimate of the variance is of the form

$$\mathrm{D}k_{\mathrm{eff}} \approx \left(\frac{\hat{n} - n_s}{\hat{n} - n_l}\right)^2 \left[\frac{1}{\hat{n} - n_s} - \frac{1}{\bar{\nu}(\hat{n} - n_l)}\right]$$

where $\bar{\nu}$ is the average number of neutrons per fission.

25. The Method of Moments

This method involves the construction of a finite matrix which approximates to the integral operator (3.11), and the determination of its maximum eigenvalue k, which is approximately equal to k_{eff}.

We shall now explain how an integral operator may be approximated by a matrix.

Suppose we have a complete set of functions $f_1(r), f_2(r), \ldots$. Any function $n(r)$ can then be expanded in terms of these functions:

$$n(r) = c_1 f_1(r) + c_2 f_2(r) + \ldots$$

Thus, to each function $n(r)$ there corresponds an infinite vector $\{c_1, c_2, \ldots\}$. The constants $c_1, c_2, \ldots$ are called the Fourier coefficients of the function $n(r)$ with respect to the complete system $f_1(r), f_2(r), \ldots$.

Suppose further that the integral operator K in

$$\tilde{n}(r) = \int K(r, r') n(r') dr'$$

transforms the functions $f_i(r)$ into $\tilde{f}_i(r)$ so that

$$\tilde{f}_i(r) = \int K(r, r') f_i(r') dr',$$

and let

$$\tilde{f}_i(r) = a_{i1} f_1(r) + a_{i2} f_2(r) + \ldots, \qquad i = 1, 2, \ldots$$

It is clear that the infinite matrix $\|a_{ij}\|$ is equivalent to the integral operator K since it may be used to find the vector $\{\tilde{c}_1, \tilde{c}_2, \ldots\}$ corresponding to the function $\tilde{n}(r)$ from the vector $\{c_1, c_2, \ldots\}$:

$$\tilde{c}_i = a_{1i} c_1 + a_{2i} c_2 + \ldots, \qquad i = 1, 2, \ldots$$

If only a finite number m of the coordinates is retained, so that the vector becomes $\{c_1, \ldots, c_m\}$, then one can consider the finite matrix $\|a_{ij}\|$, $1 \leqslant i, j \leqslant m$ which transforms the vector $\{c_1, \ldots, c_m\}$ into the vector $\{\tilde{c}_1, \ldots, \tilde{c}_m\}$. This is, in fact, the finite matrix which approximates the integral operator K. It transforms functions of the form

$$n(r)=c_1 f_1(r)+\ldots+c_m f_m(r)$$

into the functions

$$\tilde{n}(r)=\tilde{c}_1 f_1(r)+\ldots+\tilde{c}_m f_m(r),$$

so that

$$\tilde{c}_i=a_{1i}\cdot c_1+\ldots+a_{mi}c_m, \qquad i=1,\ 2,\ \ldots,\ m.$$

Gel'fand et al. [19], who were the first to use the method of moments, have calculated a matrix which approximates the operator conjugate to (3.11) (the eigenvalues of conjugate operators are equal). Consider the history of a neutron born at the point $r_0^{(s)}$. Let $\nu\left(r_{n_j}^{(s)}\right)$ be the number of secondary neutrons produced by the latter neutron at the end of the jth history at the point $r_{n_j}^{(s)}$ (in the case of escape or capture $\nu=0$). Having accumulated N such histories beginning at the same point $r_0^{(s)}$ we find the approximate values of the integrals

$$\tilde{f}_i\left(r_0^{(s)}\right)=\int g\left(r,\ r_0^{(s)}\right) f_i(r)\,dr\approx\frac{1}{N}\sum_{j=1}^{N}\nu\left(r_{n_j}^{(s)}\right) f_i\left(r_{n_j}^{(s)}\right). \qquad (3.12)$$

They are evaluated simultaneously for all $i=1,2,\ldots,m$. Having selected p such points $r_0^{(s)}$ $(s=1,2,\ldots,p)$ where $p\gg m$, and having calculated for each of them the value of $\tilde{f}_i\left(r_0^{(s)}\right)$, we choose the matrix a_{ij} so that

$$\tilde{f}_i\left(r_0^{(s)}\right)\approx\sum_{j=1}^{m} a_{ij} f_j\left(r_0^{(s)}\right), \qquad \begin{array}{l} i=1,\ 2,\ \ldots,\ m; \\ s=1,\ 2,\ \ldots,\ p. \end{array}$$

The determination of a_{ij} may be carried out, for example, by the method of least squares, so that

$$\sum_{s=1}^{p}\sum_{i=1}^{m}\frac{1}{D_{is}}\left[\tilde{f}_i\left(r_0^{(s)}\right)-\sum_{j=1}^{m} a_{ij} f_j\left(r_0^{(s)}\right)\right]^2=\min.$$

The variances D_{is} are estimated from the same case histories. The largest eigenvalue of $\|a_{ij}\|$ is then formed by purely algebraic methods.

26. Normalization of the Number of Neutrons

The following difficulty may be encountered in the realization of the methods described in paragraphs 23 and 24. When k_{eff} is appreciably greater than unity, the number of neutrons will rapidly increase and may overload the computer store before the density profile is established. Conversely, if k_{eff} is much less than unity, the number of neutrons may decrease rapidly, or may even become zero.

This degeneracy is not catastrophic when it is required to determine the value of some reactor parameter which will ensure criticality. It indicates the direction in which the parameter should be varied. However, the possibility of interpolation between the calculated values of k_{eff} is then lost. Bouquet et al. [84], using the method of paragraph 24, avoided the above degeneracy by introducing a spatial grid which enabled them to normalize the number of neutrons in each generation. The active zone was divided into elements (0.1 of a free path in diameter), and all the neutrons produced in a given element were placed at its center. The initial generation density $n_1(r)$ was replaced by the numbers of neutrons produced in these elements $n_{11}, n_{12}, \ldots, n_{1s}$ so that $n_{11}+n_{12}+\ldots+n_{1s}=M$. When the number of secondary neutrons $n_{21}, n_{22}, \ldots, n_{2s}$ had been found, the required ratio was calculated from

$$\frac{n_{21}+n_{22}+\ldots+n_{2s}}{M}.$$

This was then followed by the normalization of the neutrons: the new n_{2j} were found by multiplying the old n_{2j} by the reciprocal of the above ratio, i.e., $M/(n_{21}+\ldots+n_{2s})$.

In this method of calculation the numbers n_{ij} turn out to be fractions and play the role of weights (masses). It is possible to construct a single trajectory for each of the elements and multiply the weights of the next-generation neutrons by n_{ij}. For those elements in which $m-1 \leqslant tn_{ij} < m$, it is more convenient to initiate m histories and multiply the weights of the resulting neutrons by n_{ij}/m, where m is an integer and the value of t is chosen in accordance with the value of M.

27. Application of Weights

The advantages of the method of counting of generations, as compared with the simulation of physical trajectories, lie in the fact that ordinary trajectories can be considered instead of branching trajectories. It is thus possible to use the various methods of introducing statistical weights which were discussed above. In particular, the weights allowing for absorption (paragraph 14) and the weights discussed at the end of paragraph 26 are often employed. The statistical weights discussed in paragraph 5, which represent the escape of neutrons, are less frequently used.

More complicated formulas [analogous to (3.9)] can be used simultaneously with the statistical weights. For example, a simple count of the descendants of a neutron, i.e., the number of neutrons available at the end of a case history, may be replaced by

$$\bar{\nu} v_n \frac{\Sigma_f(E_n)}{\Sigma_t(E_n)},$$

where v_n is the statistical weight of a neutron approaching the nth collision. Similarly, the integrals in (3.12) may be estimated with the aid of the formula

$$\tilde{f}_i(r_0^{(s)}) \approx \frac{\bar{\nu}}{N} \sum_{j=1}^{N} \sum_{n=1}^{n_j} v_n \frac{\Sigma_f(E_n)}{\Sigma_t(E_n)} f_i(r_n^{(s)}).$$

in which $w_0 = 1$.

28. Determination of the Critical Values of Reactor Parameters

The methods of paragraphs 23-25 may be used to calculate k_{eff} for a given region R. However, a reactor model frequently contains some parameter $\varkappa$ whose critical value must be determined. This parameter may be, for example, the concentration of uranium in the active zone, the thickness of the reflector, and so on.

The critical value of $\varkappa$ is the value which corresponds to $k_{\text{eff}} = 1$. Thus, in order to determine $\varkappa_{\text{crit}}$ it is necessary to determine a number of values of the function $k_{\text{eff}} = F(\varkappa)$ and carry out an interpolation leading to $F(\varkappa_{\text{crit}}) = 1$.

As a rule, the same histories may be used to determine several values of the functions $F(\varkappa)$. For example, let $\varkappa$ be the thickness of the reflector, and consider a number of values $\varkappa = \varkappa_1 < \varkappa_2 < \ldots < \varkappa_r$ (Fig. 26). During the determination of the history let us record the maximum layer number reached by the neutron. If the neutron has reached layer number j and then returned to the active zone giving rise to a fission, then the fission neutrons are not included in the next generation for $\varkappa = \varkappa_1, \ldots, \varkappa_{j-1}$. This method is completely analogous to that described in paragraph 18.

29. Calculation of the Criticality on the One-Velocity Approximation

The so-called one-velocity or one-group approximation is occasionally used, mainly with fast neutron reactors. On this approximation the energy of all the neutrons is assumed to be constant. We shall consider a further method of calculating the

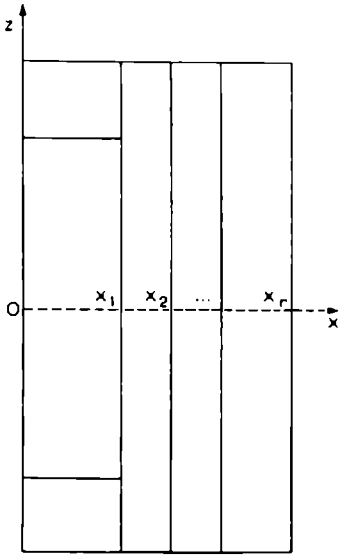


Fig. 26.

critical parameters within the framework of the one-velocity approximation (the methods of paragraphs 23-25 will still hold).

In the present case the criticality condition may be given in the form of an integral equation which is usually known as the Peierls equation. In the case of isotropic scattering this equation reads

$$\bar{n}(r)=\lambda\int_R \beta(r')K(r,\ r')\bar{n}(r')\,dr', \tag{3.13}$$

where

$$\beta=\Sigma_s+\bar{\nu}\Sigma_f,\quad \alpha=\Sigma_t,$$
$$K(r,\ r')=e^{-\left|\int\alpha\,ds\right|}\cdot|r-r'|^{-2},$$

and the integral $\int\alpha\,ds$ is evaluated over the straight line connecting the points r and r':

$$\int\alpha\,ds\equiv\int_0^{|r'-r|}\alpha(r+\omega s)\,ds,\quad \omega=\frac{r'-r}{|r'-r|}.$$

The criticality condition is of the form $\lambda_1=1$ where λ_1 is the smallest eigenvalue of the integral equation (3.13), i.e., $1/\lambda_1$ is the largest eigenvalue of the integral operator βK. The region R is subcritical or supercritical when $\lambda_1>1$ or $\lambda_1<1$, respectively.

Thus, the determination of the criticality in the one-velocity approximation is reduced to the determination of the smallest eigenvalue of the integral equation (3.13).

However, if the geometrical structure of R is complicated, then it is very difficult to find λ_1 by ordinary numerical methods.

Vladimirov [16] has suggested the determination of λ_1 by the Monte Carlo method. Such calculations have been successfully completed by Vladimirov and Sobol' [18, 59]. Since the methods employed were concerned with the solution of arbitrary integral equations, we shall not discuss them in detail. We shall merely note that the law for the construction of random trajectories which was realized by Vladimirov [18] is equivalent to the use of statistical weights of the form

$$w_n = w_{n-1}\left[1 - e^{-\int\limits_0^{\overset{*}{s}_{n-1}} \alpha(r_{n-1}+\omega_{n-1}s)\,ds}\right]\frac{\beta(r_n)}{\alpha(r_n)}$$

which represent both the escape of neutrons and the probabilistic method of determination of the interactions.*

Kellogg's method of successive approximations was used in the above papers to calculate λ_1. A variational method can also be used. In either case, it is necessary to evaluate certain complicated integrals and this can be done by the Monte Carlo method.

---

*Fission and scattered neutrons are equivalent since their energies are equal and the scattering is assumed to be isotropic.

Chapter IV

Application of the Monte Carlo Method to the Investigation of Mass Service or Congestion Processes, Including Queueing

1. GENERAL INFORMATION ON MASS-SERVICE PROBLEMS

In the course of the last decade practical needs have stimulated intensive study of a specific class of problems concerning mass service. A typical feature of such problems, which arise most often in particle physics, telephony, production scheduling, planning, automatic control of complex assemblies, etc., is the presence of a server system with which requests for service (calls) are placed at random instants of time. The server system has lines (channels) accomplishing the set of operations implied in the word "service." An automatic refueling station can be considered as an example of a server system. Requests for service (calls) arise when cars arrive at the station for refueling. The lines are the individual gasoline pumps by means of which attendants carry out the refueling of cars, ensuring independently the full cycle of operations connected with serving a call.

The object of mass-service theory is the study of the time-patterns arising in the process of serving a stream of calls forming an input in the system. This input stream is a sequence of calls with definite patterns of arrival times.

If all calls of a given stream have, from the point of view of service, equal rights, the only question which arises at a given instant of time is whether the arrival of a call takes place or not. Such streams, which are called streams of homogeneous events, have been thoroughly studied and admit of elegant and convenient mathematical description.

The solution of practical problems sometimes requires that the inhomogeneity of the stream of events be taken into account. The use of streams of homogeneous events as mathematical models to describe the real stream of calls can prove in such cases inadequate. Some such examples will be specially discussed below.

The general ideas and the methods of mass-service theory are used to formulate and to solve important applied problems. For this reason much work has been devoted to obtaining analytical relationships to enable service-performance indices to be evaluated on the basis of known parameters of the stream of calls and the characteristics of the server system.

However, the relevant results obtained in the literature concern mainly comparatively simple cases, both with respect to the structure of the stream of calls and with respect to the properties of the server. Therefore, the analytical apparatus available today cannot fully satisfy the growing demands of practice.

Investigations have shown that many problems connected with mass service can be solved by having recourse to the Monte Carlo method.

A comparison of the potentialities of the analytical methods known today in the theory of mass service and the Monte Carlo method from the point of view of evaluating service-performance indices suggests the following considerations.

Formulas and equations for evaluating service-performance indices (the probability of loss, the expectation of queue length, etc.) are, as a rule, of an asymptotic nature and are valid for instants of time sufficiently far removed from the beginning of service. Such solutions do not always prove satisfactory from a practical point of view. The fact is that real mass-service processes exhibit transient behavior from the state with full availability of the lines to the stationary state described by the asymptotic formulas. The duration of the transient state is far from always negligibly small and is sometimes comparable with the duration of the service cycle typical of the system studied.

The Monte Carlo method enables us to obtain estimates for the service-performance indices for arbitrary intervals of time, including intervals during which the behavior is transient.

Existing analytical methods for solving problems of mass-service theory proceed from very restrictive assumptions on the distribution laws of the streams of homogeneous events representative of the stream of calls: the great majority of results concern the simple stream (the Poisson stream). The Monte Carlo method enables us to obtain an effective solution of the problem under materially wider assumptions on the nature of the stream of calls. The so-called streams with limited after-effect, which are defined below, prove particularly convenient in this respect.

Similar remarks can be made in connection with the structure of the server system. The class of such systems that lend themselves to investigation by the Monte Carlo method is substantially more general than the set of systems encompassed by known analytical methods.

In comparison with analytical methods, the Monte Carlo method enables us to characterize with greater completeness the dependence of service-performance indices on the parameters of the stream of calls and of the server system. This is explained by the fact that in solving problems in mass-service theory by the Monte Carlo method we can use more extensive information on the process than is usually possible when analytical methods are employed. For example, for systems with loss the usual formulas of mass-service theory give only the mean value of the fraction of losses. In this case, the Monte Carlo method enables us to obtain an estimate not only of the mean fraction of losses, but also of any parameter of the distribution law of this fraction.

Multiphase systems and multiphase mass-service processes play an important role in practice. To this date no satisfactory analytical methods exist for investigating multiphase systems. The Monte Carlo method enables us to obtain the solution of problems for multiphase systems under very general assumptions on their structure.

The essence of the Monte Carlo method in its applications to mass-service problems consists in the following. By means of special algorithms, samples of the assigned streams of homogeneous events are generated and models of the functioning processes of the server system are constructed. These algorithms are used for repeatedly reproducing samples of the random service process under fixed conditions. Information on the process, accumulated by reproducing a set of samples, is subjected to statistical processing for estimating service performance indices. This chapter gives a brief account of those features of streams of calls and server systems which are needed for the solution of problems by the Monte Carlo method; in addition, methods of generating samples of random streams of calls and methods of constructing algorithms to simulate the service process are considered.

2. MATHEMATICAL DESCRIPTION OF AN INPUT CONSISTING OF A STREAM OF CALLS REQUIRING SERVICE

We shall begin our analysis of the problem of the mathematical description of a stream of calls with the simplest case in which the stream can be assumed to be a stream of homogeneous events.

Each event of the stream is characterized in this case by the instant of time t_i at which it occurs. To describe a stream of homogeneous events as a random process it suffices to assign the distribution law characterizing the sequence of random quantities $t_1, t_2, \ldots, t_m, \ldots$

It is often advantageous to consider, instead of the quantities $t_1, t_2, \ldots, t_m, \ldots$, the random quantities $\xi_1, \xi_2, \ldots, \xi_m, \ldots$, which are the durations of the time intervals between the successive instants of time t_i,

$$\left.\begin{aligned} t_1 &= \xi_1, \\ t_2 &= \xi_1 + \xi_2, \\ &\cdots\cdots\cdots\cdots \\ t_k &= \xi_1 + \xi_2 + \cdots + \xi_k. \end{aligned}\right\} \tag{4.1}$$

Below, we shall consider only continuous random quantities ξ_i and shall characterize their occurrence by means of the joint probability density function $f(z_1; z_2; \ldots; z_k)$.

Certain special classes of streams of homogeneous events play an important role in the applications. One is the class of streams with limited after-effect.

A random stream of homogeneous events is called a stream with limited after-effect if the random quantities ξ_i are independent.

It is evident that for streams with limited after-effect the joint probability-density function $f(z_1, z_2, \ldots, z_k)$ can be written in the form

$$f(z_1, z_2, \ldots, z_k) = f_1(z_1) f_2(z_2) \ldots f_k(z_k). \tag{4.2}$$

The functions $f_i(z_i)$ $(i > 1)$ are conditional probability density functions of the quantities ξ_i with the condition that at the initial instant of the interval ξ_i $(i > 1)$ a call has arrived. In contrast to this, the function $f_1(z_1)$ is an unconditional probability density function, since no assumptions whatsoever are made as to whether at the initial instant of time a call has or has not arrived.

Of great theoretical and practical interest are the so-called stationary streams, for which the probability structure does not depend on time.

More precisely, a stream of homogeneous events is called stationary if the probability of the occurrence of k events during the interval of time $(t_0, t_0 + t)$ is independent of t_0 and depends only on t and k.

Stationary streams with limited after-effect satisfy the relation

$$f_2(z)=f_3(z)=\ \dots\ =f_k(z)=f(z). \tag{4.3}$$

This means that, for $i>1$, the intervals ξ_i are identically distributed. An input consisting of such a stream is sometimes described as a general recurrent or general independent input.

Let us consider the expectation M of the random quantity ξ_i for $i>1$

$$\mathsf{M}=\int_0^\infty zf(z)\,dz.$$

The quantity M is the mean duration of the interval between successive calls.

It can easily be seen that for recurrent streams, i.e., stationary streams with limited after-effect, the quantity

$$\lambda=\frac{1}{\mathsf{M}} \tag{4.4}$$

is the mean number of events occurring per unit time. The parameter λ is called the stream density.

As an example of a recurrent stream consider the stream that has uniform distribution of the intervals of time between calls. In this case, the probability density function $f(z)$ has the form

$$f(z)=\frac{1}{b}\qquad(0\leqslant z\leqslant b). \tag{4.5}$$

Since the expectation of the quantity ξ is equal to $b/2$, then the stream density of the stream assigned by the probability density function (4.5) is equal to

$$\lambda=\frac{2}{b}. \tag{4.6}$$

Another example of a recurrent stream is the so-called simple stream (the Poisson stream or stream without after-effect). For a simple stream we have

$$f(z)=\lambda e^{-\lambda z}. \tag{4.7}$$

The parameter λ is the density of the simple stream.

Recurrent streams satisfy the following relation (Palm's formula, see [10]) connecting the probability density functions $f_1(z_1)$ and $f(z)$:

$$f_1(z_1) = \lambda\left[1 - \int_0^{z_1} f(u)\,du\right]. \tag{4.8}$$

By using (4.8) we can obtain the probability density function $f_1(z_1)$ for various recurrent streams. For example, for a stream with uniform distribution of the intervals

$$f_1(z_1) = \frac{2}{b}\left[1 - \int_0^{z_1} \frac{du}{b}\right]$$

or

$$f_1(z_1) = \frac{2(b - z_1)}{b^2} = \lambda\left(1 - \frac{\lambda}{2} z_1\right). \tag{4.9}$$

For a Poisson stream we have

$$f_1(z_1) = \lambda\left[1 - \lambda\int_0^{z_1} e^{-\lambda u}\,du\right]$$

or

$$f_1(z_1) = \lambda e^{-\lambda z_1}. \tag{4.10}$$

It follows from (4.10) that in the case of a Poisson stream we have $f_1(z_1) = f(z)$. As can be easily seen from the previous example, this is not true in general.

We observe that the distribution law of a stationary stream can be assigned in the form of a distribution law $P(k, t)$ of the number k of events occurring during an interval of time of duration t. For a Poisson stream $P(k, t)$ is equal to

$$P(k,\ t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{4.11}$$

i.e., it has the form of the Poisson law with parameter λt.

We shall consider, as an example of a nonstationary stream, a Poisson stream with variable parameter. The distribution law $P(k,\ t_0,\ t)$ of the number of events occurring during the interval of time $(t_0,\ t_0 + t)$ has for such a type of stream the form

$$P(k,\ t_0,\ t) = \frac{[\Lambda(t_0,\ t)]^k}{k!} e^{-\Lambda(t_0,\ t)}. \tag{4.12}$$

The quantity $\Lambda(t_0, t)$ is the expectation of the number of events occurring during the interval of time (t_0, t_0+t). The function $\Lambda(t_0, t)$ can be represented in the form

$$\Lambda(t_0, t)=\int_{t_0}^{t_0+t} \lambda(u)\,du. \tag{4.13}$$

The quantity $\lambda(t_0)$ is the instantaneous stream density at the instant t_0 and $\Lambda(t_0, t)/t$ is the mean stream density in the interval of time (t_0, t_0+t).

It can be shown that the probability density function $f_1(z_1)$ for the first interval ξ_1 has the form

$$f_1(z_1)=\Lambda_z'(0, z_1)\,e^{-\Lambda(0, z_1)}, \tag{4.14}$$

and, for an arbitrary interval, the form

$$\varphi(t, z)=\Lambda_z'(t, z)\,e^{-\Lambda(t, z)}. \tag{4.15}$$

We have considered thus far so-called ordinary streams of homogeneous events. A stream is called ordinary if the probability $\psi(t_0, t)$ of the occurrence of two or more events during an interval of time (t_0, t_0+t) for arbitrary t_0 is a small quantity in comparison with t, i.e., if

$$\lim_{t\to 0}\frac{\psi(t_0, t)}{t}=0.$$

In applications that are connected with servicing grouped calls we sometimes encounter problems which give rise to clusters of events, i.e., nonordinary streams of homogeneous events have sometimes to be considered. In order to describe nonordinary streams, we must assign, in addition to the instants t_i, the distribution of the number of calls arriving at each of the instants of time t_i. In the particular case when the number of events occurring is a random quantity independent of the instants t_i, it suffices to assign the probability p_k that at an arbitrary instant of time t_i exactly k events occur.

In solving certain very important applied problems the real streams of calls cannot be reduced unconditionally to streams of homogeneous events. To obtain the solution of such problems to an accuracy sufficient in practice, the inhomogeneity of the calls of the stream cannot be ignored.

For example, if we consider the operation of a fleet of motor vehicles as a mass-service transport system, we must take into account those properties of the calls that substantially affect transport operating conditions. Such properties may comprise (in addition to the required initial instant of service): the coordinates of the loading and unloading posts, which determine the duration of the idle and the productive runs; the load characteristics (with respect to weight, size, etc.), which determine the types and number of vehicles required for serving calls; the characteristics of delivery terms, etc.

Similar features are also of importance when other real processes connected with mass service are considered.

Therefore, each call must be characterized in the general case both by the instant of its arrival in the server system, t, and by a series of parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$. The modern apparatus of probability theory provides the means for a mathematical description of random streams of calls in this general case.

However, this is not required by the concrete problems that are considered in the present chapter. Therefore, we shall restrict outselves to such particular cases as can be met in the solution of a number of practical problems.

We shall represent the random stream of the instants t_i of the occurrence of calls in the server system as a stream of homogeneous events. The remaining random parameters of the calls $\alpha_1, \alpha_2, \ldots, \alpha_k$ can be assigned by means of a conditional multivariate distribution law $\varphi(\alpha_1, \alpha_2, \ldots, \alpha_k/t_i)$.

Since such a method of assigning a stream of calls is comparatively complicated, it is advantageous sometimes to proceed to further simplifications. These simplifications usually consist in using for the instants t_i streams of homogeneous events with limited after-effect, and assuming that some of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ are independent of t_i and of each other. In spite of these important restrictions, such methods of assigning a stream of calls prove suitable for the solution of many practical problems.

3. MASS-SERVICE SYSTEMS

The solution of practical problems requires an investigation of the processes of functioning of various systems and aggregates connected with the mass servicing of calls. It is evident that all the variety of features characteristic of real processes cannot be taken into account in their mathematical description. We need only to take into account the basic factors that determine the course of the process, and can neglect second-order effects that are unimportant from the point of view of the problem formulated. In other words, the mathematical description of the process of functioning of a real system must be based on some idealized scheme reproducing the basic features of the real process.

A mass-service system consists in general of n lines capable of simultaneously and independently serving the calls. At an arbitrary instant of time a line is found in one of two states, either free or busy.

Let us assume that at some instant of time a call arrives at the server system. If at this instant of time there are free lines then the call is admitted to service. In the opposite case, i.e., when all lines are busy, the call remains in the system for a certain time (let τ_p be the permanence time of a call in the system) as a claimant to service. Within an interval of time τ_p the call must be admitted to service, otherwise it is assumed to be lost (rejected).

Depending on the value of τ_p, mass-service systems fall into three classes differing both in the structure of the service process and in the mathematical formulation of relevant problems. If $\tau_p = 0$, a call arriving at a certain time is either immediately admitted to service (if there are free lines) or rejected (if all lines are busy). Such mass-serice systems are called loss systems. The service-performance indices usually considered for loss systems are the probability of loss, the mean number of calls lost over a given period of time, etc.

In the other limit case when $\tau_p = \infty$, calls arriving at the system are never lost, but (if all lines are busy) wait in a queue until they are admitted to service. Such mass-service systems are called delay systems. Service-performance indices considered in this case can be the mean waiting time of a call, the mean queue length, etc.

Finally, if $0 < \tau_p < \infty$, a call that finds all lines busy at the instant of its arrival waits for a time τ_p in a queue, and when this time has elapsed (without any line being made available to it) the call is rejected. Such mass-service systems are called mixed systems or combined loss and delay systems. The service-performance indices considered in this case are the probability characteristics of both the number of rejections and the waiting time as well as, sometimes, more complicated indices that take into account both these aspects of service performance.

To characterize the properties of a server system we must assign, in addition to the time τ_p, also the time τ_0, namely, the time needed to serve a call or, in other words, the service time of a line. A call admitted to service occupies one of the lines for a time τ_0; as soon as this time has elapsed the line is free and can proceed to serve another call.

The quantities τ_0 and τ_p are usually considered to be random quantities with assigned distribution laws (or else with one assigned joint distribution law). The assumption can sometimes be made that one of them or both are fixed quantities.

Let us proceed to consider various common versions of line-occupancy schedule. If in the mass-service system the incoming calls are queued, then the lines becoming free are immediately occupied in the order of their becoming free. In the case when there is no queue of calls and there are free lines, an incoming call may occupy one of the free lines according to special rules. The following rules are most often used in practice.

1. The lines are occupied according to their serial numbers. A line with a large serial number cannot be enlisted for service to a call if there is a free line with a smaller serial number.

2. The lines are occupied according to the order in which they become free. As a line becomes free it is entered in a queue and cannot be enlisted for service to a call as long as there are free lines that have become free earlier.

3. The lines are occupied in random order according to assigned probabilities.

If at the instant of arrival of the next call there are k free lines, the probabilities $p_{r_1}, p_{r_2}, \ldots, p_{r_k}$ of occupying the r_1th, r_2th, ..., r_kth line respectively must be assigned.

Similar assumptions can also be made concerning the order of admission to service for calls that are found in a queue. For example, it is natural to use the following rules or "queue disciplines":

1. Calls are admitted to service in their queueing order. As a line becomes free, it proceeds to serve the call which arrived first in the system.

2. Calls are admitted to service in the order of priority of their possible rejection. As a line becomes free, it admits to service the call that has used up the largest fractions of its τ_p and which would be rejected the soonest if it were not admitted to service. We observe that the second rule coincides with the first one when τ_p is constant.

3. Calls are admitted to service in a random order according to assigned probabilities. If at the instant of a line becoming free there are m calls in the queue, then we must know the probabilities $p_{r_1}, p_{r_2}, \ldots, p_{r_m}$ of admitting to service calls with serial numbers $r_1, r_2, \ldots, r_m$ respectively.

The description of versions of elementary idealized functioning schemes of mass-service systems can be restricted to the information above. In practice, other similar setups can also be met, which the reader can reduce to idealized form by analogy.

Let us proceed to consider distinctive features that are typical of more complicated nonelementary functioning schemes of mass-service systems.

Real mass-service processes are often multiphase, the server system consisting of a certain number of subsystems operating in tandem. As a rule, the following service discipline is usually observed: the next subsystem can proceed to serve a call only when service at the previous phase has been carried through.

We sometimes find processes characterized by the fact that all calls (including those rejected at the previous phase) are admitted to the following phase of service. As a rule, however, a call that has been rejected at one phase is eliminated from any further service.

We can consider, as an elementary example of multiphase service, the serving of customers in a shop. In the first phase, the customer is served by the assistant at the counter who demonstrates the goods and prepares the bill. In the second phase the customer pays the cashier according to the bill. Only after this does the customer proceed to the third phase, the checking-and-pickup department.

The above-mentioned rule, according to which the service operations of various phases are carried out in sequence, is not always observed in real multiphase service systems. Sometimes these operations are completely or partially simultaneous in time.

A further generalization of elementary mass-service schemes is provided by the following considerations.

When the inhomogeneity of the incoming calls is to be reckoned with, the distribution laws of the parameters τ_p and τ_0 and of other characteristics of the server system are usually found to depend on quantities $\alpha_1, \alpha_2, \ldots, \alpha_k$, describing the inhomogeneity of the calls (see the end of the previous section), or on their probability characteristics. This generalization is not one of principle, in that, strictly speaking, the case considered may be reduced to the elementary one; it does, however, substantially affect the procedure of solution of problems.

In just the same manner it may be found that the probabilities determining the order of enlisting lines and the order of admitting calls to service do not remain constant but depend on the parameters $\tau_p, \tau_0, \alpha_1, \alpha_2, \ldots, \alpha_k$ or else on their probability characteristics.

It is sometimes advantageous to consider instead of two possible states of a line—free or busy—a greater number of possible states. For example: a line is free and ready to serve calls, a line is busy, a line is free but not ready to serve calls, etc.

The possibility of wastage not connected with the occupancy conditions of the lines must be reckoned with in some mass-service systems. In this case the probability characteristics of the number of rejections or losses will depend substantially on the probability q of occurrence of wastage, which in the general case is a function of the parameters $\tau_p, \tau_0, \alpha_1, \alpha_2, \ldots, \alpha_k$, etc.

Finally, an important factor accompanying the functioning processes of many real mass-service systems is the reliability of the apparatus and equipment.

From the point of view of reliability, at a certain instant of time a line can in the simplest case either be in good working order or be in disrepair. The reliability of a line may be characterized by the probability of reliable operation $R = R(t)$, which depends on time. In certain cases, a line put out of operation quits the service process; in other cases it may be put again into service after a time τ_r. The repair time τ_r is usually assumed to be a random quantity with assigned distribution law. In general τ_r may depend on parameters of the flow and of the server system.

The fate of a call during whose service breakdown of a line occurs deserves some comment. The most common assumptions in this connection are the following: that the call is rejected; that the call remains in the system (with a total permanence time not greater than τ_p) as a claimant to service outside the queue; that the call is enlisted in the queue, etc.

We have enumerated here only the main generalizations of elementary service schemes, such as are met in investigating real processes. These generalizations enable us to encompass a series of very important practical problems. When recourse is made to the Monte Carlo method, these generalizations do not hinder the effective solution of problems but only lead to some complication in the algorithms used to simulate the system.

4. THE GENERATION OF RANDOM STREAMS OF CALLS

An essential element of the investigation of mass-service processes by the Monte Carlo method is the generation of random streams of calls. Simplicity and compactness of the algorithms used for generating streams of calls are most often the factor that decides the actual feasibility of an investigation to be undertaken.

We shall begin our account of methods of generating streams of calls with the simple case in which the calls form a stream of homogeneous events.

The joint distribution law $f(z_1, z_2, \ldots, z_k)$ of the random quantities $\xi_1, \xi_2, \ldots, \xi_k$, i.e., of the intervals of time between successive instants of occurrence of calls (4.1), will be assumed to be given. In order to obtain a sample stream of homogeneous events t_1, $t_2, \ldots, t_k$, we must generate a sample $z_1, z_2, \ldots, z_k$ of the k-dimensional random vector $\xi_1, \xi_2, \ldots, \xi_k$ and evaluate t_i according to (4.1). Methods of generating random vectors are considered in Chapter VII. As is well-known, this operation proves very cumbersome if k is large. This fact substantially restricts the use of streams of homogeneous events of general type in the solution of problems of mass-service theory by the Monte Carlo method.

The procedure of generating samples of streams of homogeneous events is considerably simplified in the case of recurrent streams.

Let a recurrent ordinary stream be assigned by the probability density $f(z)$ (4.3). The probability density $f_1(z_1)$ for the first interval z_1 will be found from (4.8). Now, according to the rules considered in Chapter VII, we can generate a random number z_1 corresponding to the probability density $f_1(z_1)$ and obtain the instant of occurrence of the first call $t_1 = z_1$. Then we generate a series of random numbers corresponding to the probability density $f(z)$, and evaluate by means of the relation (4.1) the quantities t_2, t_3, ..., t_k. The procedure described is illustrated below by a number of examples. The streams of homogeneous events considered as examples can be used as typical mathematical schemes for an approximate description of various streams met in the solution of practical problems.

Below, we shall assume to have available random numbers R_i with uniform distribution in the interval (0, 1). As is well-known (see Chapter VII), to obtain random numbers x_i with probability density $f(x)$, we must solve with respect to x_i the equation

$$\int_{-\infty}^{x_i} f(x)\,dx = R_i. \tag{4.16}$$

Relation (4.16) will be used, in addition to other methods, for generating streams of homogeneous events.

The simple stream. The probability density function $f(z)$ of the intervals between calls, ξ_i for $i > 1$, has, for the simple stream, the form

$$f(z) = \lambda e^{-\lambda z}. \tag{4.7}$$

Owing to (4.10) the same expression is valid for the probability density function $f_1(z_1)$ of the first interval ξ_1. Therefore, the generation of samples of a simple stream of homogeneous events can be reduced to the particular case of generating a sequence of independent random numbers having the exponential distribution (4.7).

To do this we use relation (4.16)

$$\lambda \int_0^{z_i} e^{-\lambda z}\,dz = R_i$$

or

$$1 - e^{-\lambda z_i} = R_i. \tag{4.17}$$

By solving equation (4.17) with respect to z_i we obtain

$$z_i = -\frac{1}{\lambda}\ln(1 - R_i). \tag{4.18}$$

Then, to obtain the corresponding sequence of the instants of arrival of calls $t_1, t_2, \ldots, t_k, \ldots$, we use (4.1).

We observe that, to find the values of z_i in accordance with (4.18), a comparatively large number of operations on general-purpose electronic digital computers must be performed. This is explained by the fact that the evaluation of the logarithms by general-purpose digital computers is carried out by means of standard programs based on power-series expansions. Therefore, to generate a sequence of random numbers with exponential distribution law, recourse is often had to approximate methods (Chapter VII).

In the particular case of a simple stream certain other methods of generating samples can also be recommended. For example, for small values of λ (in practice for $\lambda < 0.5$) the following approximate method proves advantageous. Its key idea is based on simulating the conditions of the corresponding limit theorem.

The units of time (e.g., minutes) used in this problem are given serial numbers 1, 2, ..., m, ... Let us split each unit of time into equal parts of length $\tau = \theta/\lambda$, where $0 < \theta < \lambda < 1$ and λ/θ is an integer. Within each unit of time we enumerate the intervals obtained: 1, 2, ..., j, ..., λ/θ. Then we use the following random-test procedure. From a set of random numbers with uniform distribution over the interval (0, 1) we select a random number R_i and test the inequality

$$R_i \leqslant \theta. \tag{4.19}$$

If the inequality (4.19) is satisfied, we assume that a call has arrived at the instant of time

$$t_i = (m-1) + (j-1)\tau + \frac{\tau}{\theta} R_i. \tag{4.20}$$

If the inequality (4.19) is not satisfied, we assume that no call has arrived, and proceed to the next random number R_{i+1}.

Beginning with the zero instant of time ($m=1$; $j=1$), we check, by means of the above-described tests, whether a call has arrived during the first interval (of duration τ) of the first unit of time. If the inequality (4.19) is satisfied, the instant of arrival of the first call is

$$t_1 = \frac{\tau}{\theta} R_i.$$

If the inequality (4.19) is not satisfied, we proceed to the second interval (of duration τ) of the first unit of time ($m=1$; $j=2$), etc., until all intervals of the first unit of time are exhausted. Then we proceed to the second unit of time ($m=2$; $j=1$), etc.

In order to ensure a closer compliance of the distribution law of the stream thus formed with the Poisson distribution law, we must reduce θ. An obstacle to this is the increase of the volume of computations which, for very small $\tau=\theta/\lambda$ (practically $\tau<0.05$), makes the procedure no less cumbersome than the calculation of z_i according to (4.18).

In the solution of problems connected with mass service of concrete realizations of the simple stream, the method most acceptable in practice for generating samples is usually that considered in Chapter VII, which is based on a piecewise approximation to $f(z)$.

Stream with uniform distribution of the intervals between calls. The probability density function $f(z)$ for the intervals ξ_l, for $l>1$, for such a recurrent stream has the form (4.5)

$$f(z)=\frac{1}{b} \qquad (0\leqslant z\leqslant b),$$

and the probability density function $f_1(z_1)$ of the first interval ξ_1 has the form (4.9)

$$f(z_1)=\frac{2(b-z_1)}{b^2}=\lambda\left(1-\frac{\lambda}{2}z_1\right),$$

where

$$\lambda=\frac{2}{b}.$$

We observe that the expectation of the duration of the first interval is

$$\mathrm{M}\,\xi_1=\frac{b}{3}.$$

The procedure for generating samples of a stream with uniform distribution of the intervals between calls reduces to the following.

To obtain values of the first interval z_1 we use relation (4.16)

$$\frac{2}{b^2}\int\limits_0^{z_{1j}}(b-z)\,dz=R_j.$$

Then

$$z_{1j}=b\left(1-\sqrt{1-R_j}\right), \tag{4.21}$$

where R_j are random numbers with uniform distribution over the interval (0, 1).

The expenditure of a comparatively large number of machine operations to calculate expression (4.21) is of little importance, since this is to be done very rarely, namely, once for each sample of the stream.

Random numbers having uniform distribution over the interval $(0, b)$ are used as possible values of z_i $(i > 1)$. It is evident that they can be obtained by means of the transformation

$$z_{ij} = bR_j. \tag{4.22}$$

Samples of a stream with uniform distribution of the intervals between calls are thus simply and conveniently obtained by means of general-purpose electronic digital computers.

Erlangian streams. An Erlangian stream of order m is an ordinary recurrent stream for which the probability density function of the intervals between calls ξ_i for $i > 1$ has the form

$$f(z) = \frac{\lambda^{*m}}{(m-1)!} z^{m-1} e^{-\lambda^* z}. \tag{4.23}$$

It can be shown that the density of an Erlangian stream is equal to

$$\lambda = \frac{\lambda^*}{m}. \tag{4.24}$$

The use of approximate methods considered in Chapter VII is in general recommended for obtaining random values z_{ij} of the first interval ξ_1.

The process of obtaining values z_{ij} for the intervals ξ_i for $i > 1$ may, however, be simplified if the following fact is taken into account. The intervals ξ_i for $i > 1$ of an Erlangian stream of order m are the sums of m independent components ξ_i, each of which has an exponential distribution with parameter λ^*.

Thus, the intervals z_{ij} for an Erlangian stream can be obtained by adding m sequences of values of ξ_i.

We observe that the generation of samples of an Erlangian stream by means of electronic digital machines proves rather cumbersome.

Below, we consider some streams which depend on two or three parameters and have distinct advantages over Erlangian streams from the point of view of the case of generating samples by means of electronic digital machines.

First generalization of the Erlangian stream. Let us consider an ordinary recurrent stream for which the intervals ξ_i between calls for $i > 1$ are the sums of two random quantities subject to an exponential distribution law with parameters λ_1 and λ_2.

It can be shown that the probability density function $f(z)$ for such a stream will have the form

$$f(z)=\lambda_1\lambda_2\frac{e^{-\lambda_1 z}-e^{-\lambda_2 z}}{\lambda_2-\lambda_1}. \tag{4.25}$$

The stream density is expressed by the relation

$$\lambda=\frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}. \tag{4.26}$$

The probability density function for the first interval has the form

$$f_1(z_1)=\frac{\lambda_1\lambda_2}{\lambda_2^2-\lambda_1^2}(\lambda_2 e^{-\lambda_1 z_1}-\lambda_1 e^{-\lambda_2 z_1}). \tag{4.27}$$

In order to obtain possible values z_{ij} of the intervals ξ_i for $i>1$, the procedure may be used of adding up sequences of random numbers with exponential distributions reduced to the corresponding values of λ.

In order to obtain random values z_{1j} of the first interval ξ_1, we must substitute (4.27) in (4.16). From this we obtain the relation

$$1+\frac{\lambda_1^2}{\lambda_2^2-\lambda_1^2}e^{-\lambda_2 z_{1j}}=R_j+\frac{\lambda_2^2}{\lambda_2^2-\lambda_1^2}e^{-\lambda_1 z_{1j}}, \tag{4.28}$$

which can be solved with respect to z_{1j} by a method of successive approximations.

If the intervals between calls ξ_i, for $i>1$, are the sums of three components having exponential distribution with parameters λ_1, λ_2 and λ_3 respectively, we shall have, instead of relations (4.25), (4.26) and (4.27), the following ones:

$$f(z)=\lambda_1\lambda_2\lambda_3\left[\frac{e^{-\lambda_1 z}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}+\frac{e^{-\lambda_2 z}}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}+\frac{e^{-\lambda_3 z}}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}\right]; \tag{4.29}$$

$$\lambda=\frac{\lambda_1\lambda_2\lambda_3}{\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3}, \tag{4.30}$$

$$f_1(z_1)=\frac{\lambda_1\lambda_2\lambda_3}{\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3}\left[\frac{\lambda_2\lambda_3 e^{-\lambda_1 z_1}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}+\frac{\lambda_1\lambda_3 e^{-\lambda_2 z_1}}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}+\frac{\lambda_1\lambda_2 e^{-\lambda_3 z_1}}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}\right]. \tag{4.31}$$

Second generalization of an Erlangian stream. Let us consider a recurrent ordinary stream for which the intervals ξ_i for $i>1$ are the sums of two independent random variables having uniform distribution over the intervals $(0, b_1)$ and $(0, b_2)$, respectively.

If $b_1 < b_2$, the probability density function $f(z)$ for intervals ξ_i $(i > 1)$ has the form

$$f(z) = \begin{cases} \frac{z}{b_1 b_2} & (0 < z < b_1), \\ \frac{1}{b_2} & (b_1 < z < b_2), \\ \frac{b_1 + b_2 - z}{b_1 b_2} & (b_2 < z < b_1 + b_2). \end{cases} \tag{4.32}$$

The density of such a stream has the form

$$\lambda = \frac{2}{b_1 + b_2}. \tag{4.33}$$

The probability density function for the first interval has the form

$$f_1(z_1) = \begin{cases} \frac{2b_1 b_2 - z^2}{b_1 b_2 (b_1 + b_2)} & (0 < z \leqslant b_1), \\ \frac{2(b_2 - z) + b_1}{b_2 (b_1 + b_2)} & (b_1 < z \leqslant b_2), \\ \frac{z^2}{b_1 b_2 (b_1 + b_2)} - \frac{2(z - b_2)}{b_1 b_2} - \frac{b_1 - b_2}{b_1 b_2} & \\ & (b_2 < z \leqslant b_1 + b_2). \end{cases} \tag{4.34}$$

The procedure for obtaining possible values of z_{ij} of the intervals ξ_i for $i > 1$ reduces to adding random numbers having uniform distribution over the intervals $(0, b_1)$ and $(0, b_2)$, respectively.

To generate values of the first interval ξ_1 we can use the approximate methods considered in Chapter VII.

Stream with a fixed minimum interval. To solve applied problems we have sometimes to consider streams for which the intervals ξ_i for $i > 1$ are the sums of a constant a and a random quantity ξ.

For example, if ξ has an exponential distribution with parameter λ^*, the probability density function $f(z)$ for the interval ξ_i for $i > 1$ has the form

$$f(z) = \begin{cases} 0 & \text{for} \quad 0 < z \leqslant a \\ \lambda^* e^{-\lambda^*(z-a)} & \text{for} \quad z > a. \end{cases} \tag{4.35}$$

The density of such a stream is equal to

$$\lambda = \frac{\lambda^*}{1 + a\lambda^*}. \tag{4.36}$$

The probability density function for the first interval has the form

$$f_1(z_1)=\left\{\begin{array}{ll}\frac{\lambda^*}{1+a\lambda^*} & \text{for} \quad 0<z_1\leqslant a \\ \frac{\lambda^*}{1+a\lambda^*}e^{-\lambda^*(z_1-a)} & \text{for} \quad z_1>a.\end{array}\right\} \tag{4.37}$$

The procedure for generating values z_{ij} for $i>1$ for such a stream is evident: we need to add up the constant a and possible values x_{ij} of the random quantity ξ having an exponential distribution with parameter λ^*. The quantity λ^* can be evaluated from the relation (4.36) if the required stream density λ is assigned.

In order to obtain samples z_{1j} of the first interval ξ_1 with probability density function (4.37) we can proceed in the following manner: find the probabilities $p(0, a)$ and $p(a, \infty)$ for the value of ξ_1 to be found in the respective intervals and apply relation (4.16) to (4.37). The following expressions are obtained for the probabilities $p(0, a)$ and $p(a, \infty)$:

$$p(0, a)=\frac{a\lambda^*}{1+a\lambda^*}, \tag{4.38}$$

$$p(a, \infty)=\frac{1}{1+a\lambda^*}. \tag{4.39}$$

By using (4.16) we find

$$\frac{\lambda^*}{1+a\lambda^*}\int_0^{z_1} du=R_l \qquad (0<R_l\leqslant p(0, a)), \tag{4.40}$$

and

$$p(0, a)+\frac{\lambda^*}{1+a\lambda^*}\int_a^{z_1} e^{-\lambda^*(u-a)}\,du=R_l \qquad (p(0, a)<R_l\leqslant 1). \tag{4.41}$$

By solving Eqs. (4.40) and (4.41) with respect to z_1 we obtain

$$z_{1j}=\left(a+\frac{1}{\lambda^*}\right)R_l \tag{4.42}$$

and

$$z_{1j}=a-\frac{1}{\lambda^*}\ln\left\{1-(1+a\lambda^*)\left[R_l-p(0, a)\right]\right\}. \tag{4.43}$$

The procedure for generating values of z_{1j} consists in the following. From a set of random numbers with uniform distribution over the interval (0, 1) we select a random number R_l and compare it with $p(0, a)$. If

$$R_l \leqslant p(0, a). \tag{4.44}$$

we evaluate z_{1j} according to (4.42). If the inequality (4.44) is not satisfied, then to evaluate z_{1j} we use (4.43).

Stream with variable parameter. We shall consider one of the possible methods of generating samples for a Poisson stream with variable parameter (4.12).

To obtain possible values z_l of the intervals between calls we use relation (4.16). Let us apply it to the probability density function (4.15). Then

$$\Lambda(t, z) = -\ln(1 - R_j), \tag{4.45}$$

where R_j is a random number having a uniform distribution over the interval (0, 1).

It can easily be shown, by using the probability density function (4.14), that relation (4.45) is valid not only for z_l for $l > 1$ but also in the case of the first interval ξ_1 if we put $(t) = 0$.

By solving equation (4.45) with respect to z, the required sequence of quantities z_l can be obtained.

We shall discuss in some detail only a very simple example. Let the instantaneous stream density $\lambda(t)$ be linearly dependent on time

$$\lambda(t) = at + b. \tag{4.46}$$

Then, by virtue of (4.13),

$$\Lambda(t, z) = \frac{az^2}{2} + z(at + b), \tag{4.47}$$

and expression (4.45) assumes the form

$$\frac{az^2}{2} + z(at + b) = -\ln(1 - R_j). \tag{4.48}$$

By taking $t = 0$ for the first interval ξ_1, we obtain

$$\frac{az_1^2}{2} + bz_1 = -\ln(1 - R_j) \tag{4.49}$$

or

$$z_1 = \frac{-b + \sqrt{b^2 - 2a\ln(1 - R_j)}}{a}. \tag{4.50}$$

For the second interval ξ_2 we put $t = z_1$. Then

$$\frac{az_2^2}{2} + z_2(az_1 + b) = -\ln(1 - R_j) \tag{4.51}$$

or

$$z_2 = \frac{-(az_1 + b) + \sqrt{(az_1 + b)^2 - 2a\ln(1 - R_j)}}{a}.$$

For the interval of serial number l we put $t = \sum_{l=1}^{l-1} z_l$. Then

$$z_l = \frac{-\left(a\sum_{l=1}^{l-1} z_l + b\right) + \sqrt{\left(a\sum_{l=1}^{i=1} z_l + b\right)^2 - 2a\ln(1 - R_j)}}{a}. \tag{4.52}$$

In a similar manner we can obtain a z_i sequence also for other expressions for $\lambda(t)$ different from (4.46).

Note on nonordinary streams. If the number of calls arriving at an instant of time t_i is a random quantity independent of t_i, the following procedure may be recommended for obtaining possible values of the number of calls k. Let the probability that k calls arrive at the instant t_l be equal to p_k; $\sum_k p_k = 1$. Then the quantity k must be selected by lot according to the probabilities p_k. To do this, we select the random number R_j having uniform distribution over the interval (0, 1) and compare it with successive sums of p_k.

The number of calls is equal to k in the case when

$$\sum_{l=0}^{k-1} p_l < R_j \leqslant \sum_{l=0}^{k} p_l, \tag{4.53}$$

where $p_0 = 0$.

Note on streams of a more general nature. If calls are characterized not only by their instant of arrival t_i in the system but also by the parameters $\alpha_i^{(1)}, \alpha_i^{(2)}, \ldots, \alpha_i^{(m)}$, then the generation of samples of the stream proves more complicated.

Let the stream of instants t_l be a stream with limited aftereffect, and let the parameters $\alpha_i^{(1)}, \alpha_i^{(2)}, \ldots, \alpha_i^{(m)}$ be random quantities independent of t_i. We shall assume that some of the quantities $\alpha_i^{(k)}$ are fixed, others are independent of the remaining ones, and some are mutually dependent. With this assumption, we must assign distribution laws for the independent parameters $\alpha_i^{(k)}$ and joint distribution laws for the dependent ones.

The generation of samples of a stream of such a type consists in the following. A stream of instants t_i is generated, as a stream of homogeneous events with limited after-effect, according to the rules considered above. For each call arriving at an instant of time t_i, we choose random values of the parameters $\alpha_i^{(1)}, \alpha_i^{(2)}, \ldots, \alpha_i^{(m)}$ by using the methods for obtaining random numbers and multidimensional random vectors with assigned distribution laws considered in Chapter VII.

5. STRUCTURE OF AN ALGORITHM FOR SOLVING MASS-SERVICE PROBLEMS BY THE MONTE CARLO METHOD

The solution of mass-service problems by the Monte Carlo method in the form considered amounts essentially to simulating, by means of general-purpose digital computers, the process of serving calls arriving at the system.

General rules for the construction of simulating algorithms are not yet available. There are, however, various methods that enable us to represent a formalized service process in the form of a sequence of operations (or group of operations) accomplished by machine.

The structure of an algorithm to simulate the functioning of a typical mass-service system will be considered here by means of a concrete example.

Suppose the mass-service system consists of n lines at which calls arrive at random instants of time t_i. The calls form an ordinary recurrent stream of homogeneous events. If at the instant t_1 there are free lines, a call is admitted to service and occupies only one of the lines for a time τ_0. A call that finds all lines busy remains in the system for a time τ_p. If during this time none of the lines starts to serve the call, the call is rejected.

As lines become free, and in that order, they are occupied by calls arriving at the system; a given line cannot be occupied while there remain free lines that have become free earlier.

If at the instant of a line becoming free there are calls in the queue, the call admitted to service is the one that would be first rejected if it were not admitted to service.

Figure 27 shows the logical block diagram of an algorithm for simulating the process of serving calls of the system described above. Each operator is, as a rule, a subalgorithm which implements in the simulating process a determinate function of the

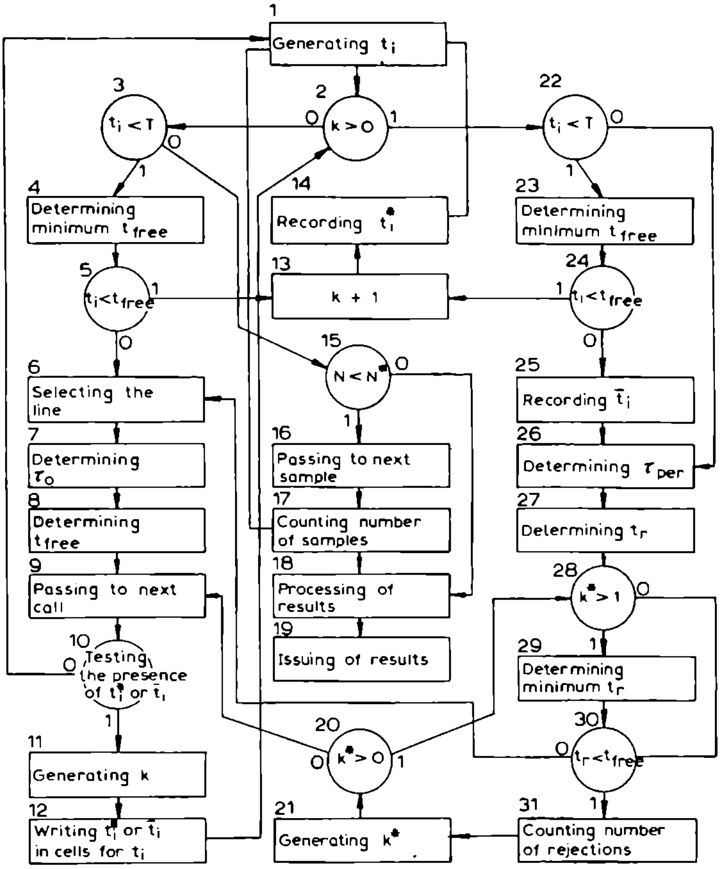


Fig. 27. Block diagram of the simulated algorithm.

system. Logical operators are shown with circles. If a condition tested by a given logical operator is fulfilled, the arrow denoting the direction of control is qualified by the index 1, in the opposite case by the index 0.

The sequence of operations in the realization of such an algorithm may be described as follows.

The operator 1 generates the instants t_i of arrival of calls in the server system. Methods of generating streams of calls have been considered in the previous section.

Suppose the operator 1 has generated the next value in turn, t_i. Control is transferred to the operator 2. By comparing with zero the value k generated by the operators 11 and 13 for the number of calls in the queue, the operator 2 produces an index equal to 1 if there is a queue, and equal to zero if there is no queue of calls.

Let us first consider the branch of the algorithm corresponding to the absence of a queue. In this case, control is transferred to the operator 3, which compares the quantity t_i with the constant T, namely, the duration of the interval of time during which the service process is to be investigated. If $t_i < T$, the call belongs to the portion of the stream that is to be investigated. In the opposite case the call is not considered.

Let $t_i < T$. Then control is transferred to the operator 4, which compares the instants at which the lines are freed (these are generated by the operator 8) and determines the smallest of them. Control is transferred to the operator 5, which establishes whether t_i is smaller than the minimum value t_{free}. If it is, this means that at the instant of the call all lines are found busy; in the opposite case there is at least one free line.

Let us first consider the case when a free line is available. Control is transferred to the operator 6, which selects, according to the assigned rule, a line for serving the given call. Then, a random value is generated for the service time τ_0 (the operator 7) as a random number with given distribution law; the time for the line to become free is calculated (to the instant t_i, which in this case is the instant of start of service, there is added τ_0; the operator 8), and the algorithm is made ready to simulate the service process of the next call (the operator 9).

The passage to the next call can be accomplished in two ways. In the first case we pass from the operator 10 along the arrow with index 0 to the operator 1 if the instant of arrival of the next call has not been generated. The second case occurs when the instant of arrival of the next call has already been generated and recorded: either during the time between subsequent instants of lines becoming free a number of calls have arrived (t_i^* is recorded by the operator 14) or the next call has a t_i that is larger than the instant at which the given line became free and therefore it is not considered as a claimant to service by the given line (t_i^* is recorded by the operator 25). If it is found that there are already values of t_i^* or $\bar{t}_i$ (this fact is verified by the operator 10), then control is transferred from the operator 10 along the arrow with index 1 to the operator 11. It is the task of the operator 11 to establish the new value of the quantity k, namely, the number of calls in the queue.

The need for this arises in connection with the fact that the values of t_i^* and $\bar{t}$ are rewritten in cells in which the instants of arrival of calls t_i (the operator 12) are usually stored. Control is then transferred to the operator 2. This concludes the analysis of the sequence of operators (beginning with the operator 2) for the case when at the instant t_i there is no queue.

Let us return to the operator 5. Let us assume now that the condition tested by the operator 5 is satisfied. This means that at the instant of arrival of the call t_i all lines have been found busy. Then the following facts must be taken into account. Although at the instant t_i there was no queue of calls, still during the interval of time between t_i and the instant of a line becoming free several more calls may arrive, and then, at the instant of a line becoming free, a queue of calls may be present. Therefore we pass from the operator 5 along the arrow with index 1 to the operator 13 which adds unity to the value of k (the number of calls in the queue); we record t_i with the notation t_i^* (the operator 14) and pass to the operator 1 for generating the next instant of arrival of a call.

Since now in the queue there is at least one call, the operation of the algorithm can be continued from the operator 2 only in the direction of the operator 22. However, before considering the chain of operators for the case of presence of a queue of calls, we shall return to the operator 3.

Let the condition $t_i < T$ tested by the operator 3 be found to be not satisfied. This means that the call arriving at the instant t_i does not belong to the portion of the stream being investigated and therefore is not considered. At the same time, this means that the given stream sample has been exhausted and we have to pass to the next sample if the given one is not the last. We pass, therefore, from the operator 3 along the arrow with index 0 to the operator 15. If $N = N^*$ (the condition tested by the operator 15 was not satisfied; the given sample proved to be the last), we pass from the operator 15 along the arrow with index 0 to the processing (the operator 18) and issuing (the operator 19) of the results of simulation.

In the case when $N < N^*$, the work of the algorithm is continued (starting from the operator 15) in the direction of the arrow with index 1. Here the algorithm is made ready to simulate the process of serving a new sample of the stream of calls (the operator 16), the number N of samples investigated is recorded (the operator 17), and then control is transferred to the operator 1.

Let us now proceed to consider the branch of the algorithm for the case when there is a queue of calls.

If $k > 0$ (the operator 2), we pass to the operator 22. Suppose that the condition tested by the operator 22 is satisfied. Then there remains to compare (the operator 24) t_i with the smallest time for a line to become free, which time is generated by the operator 23. If t_i is smaller than the time for a line to become free, it can be found that several more calls have the same property, i.e., arrive earlier than a line becomes free. Therefore we pass from the operator 24 in the direction of the arrow with the index 1 to the subchain of operators considered earlier (13, 14 and 1). The work of the algorithm will continue cyclically (the operators 1, 2, 22, 23, 24, 13, 14, 1, etc.) until the next t_l is found larger than the time for a line to be free. In this case the buildup of the queue is interrupted and the last value of t_l is recorded with the notation $\bar{t}_l$ by the operator 25. We observe that a call that has arrived at the instant $\bar{t}_l$ (after the given line has become free) is not a claimant to service by the given line. Thus, at the instant of a line becoming free there is a certain number k^* of calls in the queue which are claimants to service by the given line. The instants t_l^* of their arrival in the system are recorded by the operator 14.

The operator 26 generates, for all calls arriving at the instants t_l^*, random values of the permanence time in the system $\tau_{\text{per } l}$ as random numbers with assigned distribution law.

By adding the values of $\tau_{\text{per } l}$ to t_l^* (the operator 27) we obtain the instants of time $t_{\text{r} l}$ at which calls can be rejected if they have not been admitted to service.

The operator 28 compares the value of k^* with unity. If at the instant of a line becoming free there were found several calls in the queue ($k^* > 1$), we need to determine the smallest value $t_{\text{r} l}$ (the operator 29), since the line that becomes free must proceed to serve calls with the least value of t_{r}. If $k^* = 1$, control is transferred directly to the operator 30. The operator 30 compares the value $t_{\text{r} l}$ with the time t_{free} of a line becoming free. The permanence time in the system for calls that satisfy the condition tested by the operator 30 elapses before a line becomes free. Therefore such calls will not be served (are rejected).

The number of rejected calls (losses) is recorded by the operator 31. Then the rejected call is eliminated (the operator 21) from the number of claimants to service, and control is transferred to the operator 20. The operation of the algorithm proceeds cyclically (the operators 30, 31, 21, 20, 28, 29, 30, etc.) until either all rejected calls are eliminated from consideration and the condition tested by the operator 30 is found not to be satisfied by the following call, or else k^* becomes equal to zero.

Let $t_{\text{r}} = t_{\text{free}}$. Then the possibility of serving the call exists. In this case control is transferred from the operator 30 in the direction of the arrow with index 0 to the operator 6 and operation of the algorithm proceeds along the chain considered earlier.

It may be found, on the other hand, that $k^* = 0$ (the operator 20). This means that all calls that are claimants to service by the given line are rejected. Then along the arrow with index 0 control is transferred to the operator 9 for passing to the next call.

We observe that if the condition tested by the operator 22 is not satisfied (i.e., the call arriving at the instant t_l is eliminated from consideration) then, along the arrow with index 0, we pass to the processing of the calls that are in the queue (the operator 26 and following ones).

Algorithms for simulating call-service processes can be designed in a similar manner also for other types of mass-service systems.

The structure of a simulating algorithm for the case when calls are served in their queueing order and stoppages of the apparatus occur owing to its inadequate reliability is considered in [11].

In concluding, we observe that the same design principles can be used for the simulating algorithm for the service processes of calls with substantial inhomogeneity, when each call is characterized not only by t_i but also by a series of parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$. Typical features of these algorithms are usually the presence of operators that find the values of τ_p, τ_0, τ_{free}, τ_r, etc., as a function of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$, as well as more complicated rules for determining the order of serving calls and the order of succession of the lines used.

Algorithms to simulate call-service processes are implemented fairly efficiently by modern general-purpose digital computers.

6. CONSIDERATIONS ON THE PROCESSING OF THE RESULTS OF SIMULATION

The Monte Carlo method enables us to carry out sufficiently thorough investigations of mass-service processes of calls of a given stream. In formulating a problem in the study of mass-service processes, we must specify the quantities to be used as service-performance indices, since the structure of the simulating algorithm and the method of solution of problems often depend on our choice of the quantities to be evaluated.

For example, for systems with losses, parameters of the distribution law of the fraction of losses are usually employed as service-performance indices. The simplest and most useful index of such a kind is the mean value of the fraction of losses $R(t_0, t)$ over intervals of time (t_0, t_0+t). We shall discuss briefly the meaning of this quantity. Let us consider a set of samples of the service process over an interval of time (t_0, t_0+t). The number of calls arriving in the system during this interval of time for a sample chosen at random is a random quantity.

Let the mean value of this quantity be $N(t_0, t)$. We shall denote by $m(t_0, t)$ the mean number of losses during this same interval of time. Then

$$R(t_0, t) = \frac{m(t_0, t)}{N(t_0, t)}. \tag{4.54}$$

In the case of a stationary stream of calls the quantity $N(t_0, t)$ does not depend on t_0 and can be written in the form

$$N = \lambda t, \tag{4.55}$$

where λ is the density of the stream of calls.

For mass-service systems with constant parameters and for instants of time sufficiently far removed from the beginning of service, the quantity $m(t_0, t)$ will also be independent of t_0 and can be written as

$$m = \lambda_{\mathrm{rej}} t, \tag{4.56}$$

where λ_{rej} is the density of the stream of lost calls.

Then the mean fraction of losses R is a constant quantity

$$R = \frac{m}{N} = \frac{\lambda_{\mathrm{rej}}}{\lambda}. \tag{4.57}$$

We observed that in the case when relation (4.57) is valid, the quantity R can also be interpreted as the probability of loss for a call arriving in the system at an arbitrary instant of time.

The service-performance indices usually considered for delay systems are the mean value of the waiting time or the mean value of queue length. For mixed mass-service systems both groups of quantities are usually employed.

Estimates for the service-performance indices can be obtained from results of repeated simulation.

Let us consider, for example, the estimation of parameters connected with the distribution law of the number of losses. We shall assume that in the simulating process the number of investigated samples N is counted and that, in addition, the number of arriving calls n_i and the number of lost calls are recorded for each sample.

The last information can be used to calculate the number of samples $m_0, m_1, m_2, \ldots, m_k$ for which $0, 1, 2, \ldots, k$ losses have been recorded respectively. It is evident that we can take as estimates for the probabilities p_i, which assign the distribution law of the number of losses, the quantities

$$p_i \approx \frac{m_i}{N}. \tag{4.58}$$

The service-performance indices usually adopted can be expressed in terms of these quantities. Thus, for example, the fraction of stream samples for which there has been found at least one loss has the form

$$p \approx 1 - \frac{m_0}{N}. \tag{4.59}$$

The mean value of the number of losses occurring in a single sample can be evaluated at

$$m \approx \sum_{l} l \frac{m_l}{N}, \tag{4.60}$$

where the summation is carried out over all values of l.

The estimated mean number of calls arriving in the system during one sample will be

$$n \approx \frac{1}{N} \sum_{l=1}^{N} n_l. \tag{4.61}$$

Finally, the mean fraction of losses is equal to

$$R \approx \frac{m}{n}. \tag{4.62}$$

When the only service-performance index of interest is the mean fraction of losses, the procedure of recording and processing of the results of simulation may be simplified. In fact, we shall record only the number of calls n^* arriving in the system during all N samples of the stream, and the number m^* of those of them that are lost. Then

$$R \approx \frac{m^*}{n^*}. \tag{4.63}$$

Chapter V

Application of the Monte Carlo Method to Information Theory

The development of communications, telemetry, radar and other branches of engineering has raised the need for a theoretical investigation of statistical phenomena connected with the reception and deciphering of transmitted information. As a result, a new branch of science has arisen, variously referred to as "statistical communication theory," "theory of random signals and noise," etc. We shall refer to it as information theory.

Problems of information theory form one of the important fields of application of the Monte Carlo method. We have selected for discussion in this chapter typical problems of information theory that require the use of this method. No systematic account of information theory will be given here.

Attention has been devoted mainly to the analytical aspect of the problem, since a simple examination of the expressions (for example, integrals) which we encountered in this approach convinces us that classical methods are unsuitable and suggests the use of the Monte Carlo method.

The greater part of this chapter is devoted to evaluating multiple integrals encountered in an important branch of information theory, namely detection theory.

In the last section of the chapter we consider methods of simulating certain information transmission processes that do not lend themselves to investigation by known analytical methods.

1. STATISTICAL PROPERTIES OF SIGNALS AND NOISE

Communication systems of all types can be reduced to the single typical diagram shown in Fig. 28. The transmission of useful information* is accomplished in engineering by means of electromagnetic, acoustical, and other means. Whatever the chosen physical carriers of information, at the transmitting end of the communication line there are generated suitable oscillations (electromagnetic, acoustical, etc.) which are transmitted along the communication channel.* The types of oscillations used are very varied, from the simplest monochromatic oscillations to complex ones having arbitrarily wide and complicated spectra.

---

*By this term we mean not only explicit messages but also any information of interest to the recipient, for example, the very fact of the presence of a signal.

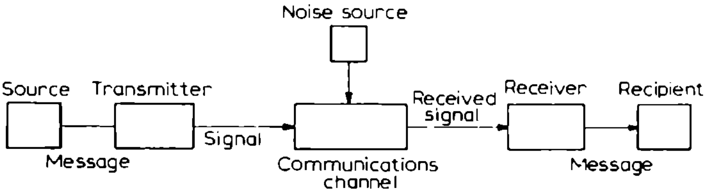


Fig. 28. Diagram of communication system.

As it passes through the communication channel the useful signal is distorted by noise, while as it passes through the receiver the fluctuation noise of the amplifier is superimposed on it. The information contained in the signal received is assessed at the output of the receiver by measuring the values of one or several physical quantities.

The simplest and most often used method of signal transmission is the excitation of monochromatic oscillations, in which case the quantity transmitted s depends on time in the following manner:

$$s(t) = A\cos(\lambda t + \varphi), \tag{5.1}$$

where A is the signal amplitude, λ is the frequency and φ is the initial phase. The transmission of such a signal enables the recipient of the message to measure, in the best case, three parameters. Therefore, in order to transmit various messages, one or several of these parameters have to be varied with time (the signal must be "modulated"). By observing the behavior of this process, the recipient must be able to extract the message transmitted. We shall consider only the method of discrete modulation, in which the information transmitted by the signal is encoded in the form of successive packets of sinusoidal oscillations differing from each other in any given parameter, for example, in amplitude. Each of such packets will be termed a signal element, and the signal itself will be considered as a time sequence of such elements. Physical elements are interpreted as variations of electrical voltage. Numerical values can be assigned to a sinusoidal pulse by describing any or all of four variables: duration, amplitude, frequency and initial phase. Below, the pulse duration will not be taken into account since its rôle in many problems is a minor one.

---

*The communication channel can be any medium in which oscillations carrying the useful information are propagated.

A monochromatic oscillation can also be represented in the following modified form of (5.1)

$$s(t) = x \cos \lambda t + y \sin \lambda t, \qquad t_1 \leqslant t \leqslant t_2,$$

where $x = A \cos \varphi$ and $y = A \sin \varphi$ are the projections of the function $s(t)$ on the orthogonal functions $\cos \lambda t$ and $\sin \lambda t$. Just as the amplitude A and the phase φ, the quantities x and y determine uniquely the form of the oscillation. When two oscillations of one and the same frequency are combined the corresponding x and y projections add up algebraically; when (5.1) is multiplied by a number the projections are also multiplied by this number. It follows from this that the projections x and y can be considered as the components of a vector describing the given oscillation.* Thus each signal element is a two-dimensional vector while a signal with n elements is described by a set of n two-dimensional vectors.

Let us denote the signal transmitted by S, its elements by s_1, $s_2, \ldots, s_n$, and the cosine and sine projections of the kth element by s'_k and s''_k respectively. Then the signal S can be written either as a vector

$$S = (s_1, s_2, \ldots, s_n),$$

or as a rectangular matrix

$$S = \begin{pmatrix} s'_1 & s'_2 & \ldots & s'_n \\ s''_1 & s''_2 & \ldots & s''_n \end{pmatrix}.$$

Wide use is made in engineering of more complicated processes. The most important among these are the following: amplitude-modulation processes, in which the amplitude is a function of time

---

*An equivalent description uses the complex notation

$$\hat{s}(t) = u e^{i\lambda t},$$

where $u = Ae^{i\varphi}$ is the complex amplitude. The real oscillatory process is characterized by the real part of the function $\hat{s}(t)$.

$$s(t) = A(t)\cos(\lambda t + \varphi),$$

phase-modulation processes, in which the phase is a function of time

$$s(t) = A\cos[\lambda t + \varphi(t)],$$

frequency-modulation processes

$$s(t) = A\cos[\lambda(t)t + \varphi],$$

as well as other processes obtained by combining the ones enumerated above. Their spectra consist no longer of a single frequency, but of a band of frequencies centered about the "carrier" frequency λ. The interpretation of the voltages of such signals as two-dimensional vectors proves rather difficult, but in the very common case of narrow-band processes, in which the modulated quantity, whether the amplitude $A(t)$ or the phase $\varphi(t)$, varies much more slowly than the oscillating function $\cos\lambda t$, such an interpretation is possible.

The information transmitted is contained in the characteristics of the signal voltages: either in the amplitude, or in the phase, or in both. The variation of these characteristics from signal element to signal element can be accomplished by various methods depending on the chosen method of exciting oscillations and on the information content. We need only point out that, in all methods, the variation of the physical characteristics is controlled by suitably assigning the values of parameters $\alpha, \beta, \gamma, \ldots$, which regulate the voltage characteristics; their concrete choice is determined by the content of the information transmitted. Therefore, a detailed notation for the signal must have the form

$$s(\alpha, \beta, \gamma, \ldots) = \begin{pmatrix} s_1'(\alpha, \beta, \gamma, \ldots) \ldots s_n'(\alpha, \beta, \gamma, \ldots) \\ s_1''(\alpha, \beta, \gamma, \ldots) \ldots s_n''(\alpha, \beta, \gamma, \ldots) \end{pmatrix}. \tag{5.2}$$

The various types of signals will be divided into two classes: deterministic and random. We shall assign to the first class signals in which the dependence of the elements on the parameters has a determinate functional form, the parameters themselves assuming fixed values. The second class consists of signals that depend stochastically on the parameters. This denotes either a random relation between elements and parameters, or randomness of the values of the parameters, or, finally, both together.

Both independently fluctuating and correlated random signals are considered. The presence of a constant component varying in the general case from element to element is admitted.

Of basic importance are the probability distributions of the combination of signal and noise, as well as those of noise alone. Among the various ways in which signal and noise combine, the one most often used is the additive combination of signal and noise. If by N we denote noise, expressed either by the vector

$$N=(n_1, n_2, \ldots, n_n)$$

with elements $n_1, n_2, \ldots, n_n$, or by the rectangular matrix

$$N=\begin{pmatrix} n_1' & n_2' & \ldots & n_n' \\ n_1'' & n_2'' & \ldots & n_n'' \end{pmatrix}, \tag{5.3}$$

where $n_1', n_1'', \ldots, n_n', n_n''$ are the components of the vectors of the "noise" voltages, we have at the output of the receiver the resultant voltage Σ, equal to the sum of the corresponding elements of signal and noise

$$\Sigma=S+N$$

or

$$\Sigma=(s_1+n_1, s_2+n_2, \ldots, s_n+n_n),$$

which in matrix form is represented thus

$$\Sigma=\begin{pmatrix} s_1'+n_1' & s_2'+n_2' & \ldots & s_n'+n_n' \\ s_1''+n_1'' & s_2''+n_2'' & \ldots & s_n''+n_n'' \end{pmatrix}. \tag{5.4}$$

It is assumed below that signal and noise are statistically independent. Expression (5.4) is conveniently written in a different form by introducing the more compact notation: $x_{2k-1}=s_k'+n_k'$ and $x_{2k}=s_k''+n_k''$, as well as $z_k=s_k+n_k=(x_{2k-1}, x_{2k})$. Thus the resultant signal will be represented in one of the forms:

$$\Sigma=(z_1, z_2, \ldots, z_n),$$

$$\Sigma=\begin{pmatrix} x_1 & x_3 & x_5 & \ldots & x_{2n-1} \\ x_2 & x_4 & x_6 & \ldots & x_{2n} \end{pmatrix}.$$

Sometimes we shall represent the signal Σ as a $2n$-dimensiona vector

$$X=(x_1, x_2, x_3, \ldots, x_{2n}). \tag{5.5}$$

Even if the signal is nonrandom, the recipient of the message has to decide whether the useful signal is present or not in the input voltage of the receiver. In the case of a positive answer, the signal parameters have to be measured as accurately as possible. This problem proves particularly difficult in the presence of a high level of noise masking the useful signal. Therefore, knowledge of the statistical properties of noise plays a particularly important role. Numerous theoretical and experimental investigations lead to the conclusion that the components of noise-voltage vectors are distributed, as a rule, according to the normal law. We shall distinguish two cases: correlated noise, when the noise elements are stochastically connected with each other, and uncorrelated noise. It is assumed, below, that the expectations of the noise components are equal to zero, i.e., that noise contains no constant components.

It is well known that the sum of multidimensional normal quantities is also a normal quantity, and that its covariance matrix is equal to the sum of the corresponding matrices of the component quantities if these are independent. Thus, when mutually independent signal and noise are combined additively we have for the $2n$-dimensional probability density function the expression

$$p_{SN}(x_1, \ldots, x_{2n}) = \frac{1}{\sqrt{(2\pi)^{2n} D}} e^{-\frac{1}{2}(X-A)\mathfrak{M}^{-1}(X-A)}, \tag{5.6}$$

where the following notation has been introduced: $\mathfrak{M} = \mathfrak{M}_S + \mathfrak{M}_N$ is the covariance matrix of the sum of signal and noise ($\mathfrak{M}_S$ and $\mathfrak{M}_N$ are the matrices of signal and noise respectively); D is the determinant of the matrix $\mathfrak{M}$; $X\mathfrak{M}^{-1}X$ is the quadratic form corresponding to the $2n$-dimensional vector X; $A = (a_1, a_2, \ldots, a_{2n})$ is a vector equal to the expectation of the vector X. The probability density of noise is represented in this notation by the formula

$$p_N(x_1, \ldots, x_{2n}) = \frac{1}{\sqrt{(2\pi)^{2n} D_N}} e^{-\frac{1}{2}X\mathfrak{M}_N^{-1}X}. \tag{5.7}$$

If the signal is nonrandom, the matrix $\mathfrak{M}$ is the covariance matrix of noise only, and the vector A represents the components of the voltages of the useful signal. Therefore, the probability density (5.6) characterizes all types of signals and noise considered. In the particular case when the components of the output voltage are independent we have

$$p_{SN}(x_1, \ldots, x_{2n}) = \frac{1}{(2\pi)^n \sigma_1 \ldots \sigma_{2n}} e^{-\frac{1}{2}\sum_{k=1}^{2n} \frac{(x_k - a_k)^2}{\sigma_k^2}}.$$

The probability density (5.6) gives an exhaustive description of the statistical properties of the output voltage and depends on all parameters of the useful signal. The mathematical investigation of detection capabilities and performance consists in the application of operations of various kinds to the normal distribution (5.6). For some widely used signals, however, the probability density (5.6) must, as a preliminary, be transformed to another functional form. This is explained by the fact that useful information is often transmitted not by means of all the characteristics of a voltage (amplitude, phase, frequency), but only by certain of them or even only one of them, for example, the amplitude. The passage to the amplitudes means analytically a change of the variables in (5.6), and is precisely the passage from Cartesian to polar coordinates with consequent averaging with respect to the polar angles (in radio-engineering language, with respect to the phases). As a result of such a transformation, the functional form of the probability density for many types of signal and noise turns out to be extremely complicated, and problems that are solved in closed form for normally distributed signals become in these cases untractable.

Let us enumerate now the modifications of the normal probability density which we shall deal with in the following exposition.

We shall begin with the simple case in which the successive elements of the message transmitted, which are formed by useful signal and additive noise, are statistically independent, and the components of each of the elements are also independent and have equal dispersions. The corresponding probability density has the form

$$p(x_1, \ldots, x_{2n}) = \prod_{k=1}^{n} \frac{1}{2\pi\sigma_k^2} e^{-\frac{x_{2k-1}^2 + x_{2k}^2}{2\sigma_k^2}}. \tag{5.8}$$

By passing to the polar coordinates

$$\left.\begin{aligned} x_{2k-1} &= r_k \cos\varphi_k, \\ x_{2k} &= r_k \sin\varphi_k, \end{aligned}\right\} k = 1, \ldots, n, \tag{5.9}$$

we obtain

$$p(r_1;\ \varphi_1;\ r_2;\ \varphi_2;\ \ldots;\ r_n;\ \varphi_n)=\begin{cases}\prod\limits_{k=1}^{n}\dfrac{r_k}{2\pi\sigma_k^2}e^{-\frac{r_k^2}{2\sigma_k^2}} & \text{for}\quad r_k>0,\\ 0, \text{ if at least one } r_k \text{ is negative.}\end{cases}$$

By averaging with respect to the phases $\varphi_1, \ldots, \varphi_n$ over intervals $(0, 2\pi)$ we arrive at the joint probability density of the amplitudes of the signal

$$p(r_1, \ldots, r_n)=\left\{\begin{array}{l}\prod\limits_{k=1}^{n}\dfrac{r_k}{\sigma_k^2}e^{-\frac{r_k^2}{2\sigma_k^2}}\ \text{for}\ r_k>0,\\ \quad 0,\ \text{if at least one } r_k\\ \qquad \text{is negative}\end{array}\right\}k=1, \ldots, n, \tag{5.10}$$

which is the product of the so-called Rayleigh unidimensional probability densities. This expression will be termed the distribution density of noiselike signal. The recipient of a signal with such a distribution assesses the information content from the amplitudes of the voltages received, the values of which are characterized by the values of the parameters σ_k^2, equal to the sum of the dispersions of the individual signal and noise voltages.

Somewhat more complicated is the expression for the probability density of the amplitudes of the voltages if the signal is rigorously deterministic. The components of the vectors of the signal voltages will be written, for the sake of simplicity, in the form

$$(s_1, 0), (s_2, 0), \ldots, (s_n, 0).$$

Then the joint probability density of the output voltages of the receiver will be equal to

$$p(x_1, \ldots, x_{2n})=\prod_{k=1}^{n}\frac{1}{2\pi\sigma_k^2}e^{-\frac{(x_{2k-1}-s_k)^2+x_{2k}^2}{2\sigma_k^2}}$$

Again, by passing to polar coordinates according to (5.9) and averaging with respect to the phases, we obtain the distribution density of the amplitudes

$$p_{SN}(r_1, \ldots, r_n) = \begin{cases} \prod\limits_{k=1}^{n} \frac{r_k}{\sigma_k^2} I_0\left(\frac{r_k s_k}{\sigma_k^2}\right) e^{-\frac{r_k^2+s_k^2}{2\sigma_k^2}} & \text{for} \quad r_k > 0, \\ 0, & \text{if at least one } r_k \text{ is negative.} \end{cases} \tag{5.11}$$

which proves to be the product of the so-called Rice probability distribution densities.* The derivation of (5.11) takes into account the fact that the signal is not random. This formula retains, however, its validity also in the case of a random signal having nonzero expectation. The value of each amplitude is characterized by two parameters s_k and σ_k, one of which, s_k, is only related to the signal, while the other one, σ_k^2, is the sum of the dispersions of useful signal and noise. The presence of a constant component in the signal (which is expressed by the fact that the expectations of the components of the vector of the signal voltage are not equal to zero) introduces complication in the form of the probability density of the amplitudes. Its analytical expression contains Bessel functions. Even more substantial complications arise in the case of statistical dependence between the signal elements.

Let us consider an example of signal with correlation between the elements of the input voltage. Let the signal elements have the form

$$S = (A_1 s, A_2 s, \ldots, A_n s),$$

where $A_1, A_2, \ldots, A_n$ are second-order matrices while s is a two-dimensional random vector with independent components. The transformations $A_k = \begin{Vmatrix} a_k & b_k \\ c_k & d_k \end{Vmatrix}$ are assumed to transform vectors with independent components into similar vectors. It can easily be verified that, for this to be so, the elements of the matrix A_k must satisfy the relations

---

*The function $I_0(x)$ is the modified Bessel function of the first kind and zero order and is defined by the relation $I_0(x) = J_0(ix)$. It is represented by the power series $I_0(x) = \sum\limits_{n=0}^{\infty} \frac{x^{2n}}{(n!)^2 2^{2n}}$. Its asymptotic behavior at infinity is described by the formula

$$I_0(x) = \frac{e^x}{\sqrt{2\pi x}}\left[1 + O\left(\frac{1}{x}\right)\right].$$

$$a_k c_k + b_k d_k = 0, \qquad k = 1, 2, \ldots, n.$$

Signals of such a form are met in some radar problems, in which

$$A_k = h_k U^k,$$

where h_k is a sequence of positive numbers and U is an orthogonal matrix. The resultant output signal is formed by the combination of the signal S with noise, the elements of which are independent. In order to evaluate the joint probability density of the amplitudes of the resultant voltage we shall fix the value of the signal-vector S. Then, owing to the independence of noise, the amplitudes of the resultant signal are found to be independent and each of them is distributed according to Rice's law. Thus the joint conditional probability density of the amplitudes, with the condition that the amplitude of the useful signal have values s, is found to be equal to

$$p(r_1, \ldots, r_n | s) = \prod_{k=1}^{n} \frac{r_k}{\sigma_k^2} e^{-\frac{r_k^2 + a_k^2 s^2}{2\sigma_k^2}} I_0\left(\frac{r_k a_k}{\sigma_k^2} s\right),$$

where $a_k = \frac{\| A_k s \|}{\| s \|}$. The unconditional joint probability density will be obtained by averaging this expression over all possible values of s. The modulus of the signal-voltage vector s is distributed, as usual, according to the Rayleigh law

$$p(s) = \frac{s}{q^2} e^{-\frac{s^2}{2q^2}}, \qquad s \geqslant 0.$$

On some reduction we obtain the following final expression of the required probability density

$$p_{SN}(r_1, \ldots, r_n) = \begin{cases} \prod\limits_{k=1}^{n} \frac{r_k}{\sigma_k^2} e^{-\frac{r_k^2}{2\sigma_k^2}} \int\limits_0^{\infty} \frac{s}{q^2} e^{-\left(\frac{1}{q^2} + \sum \frac{a_k^2}{\sigma_k^2}\right) \frac{s^2}{2}} \cdot \prod\limits_{j=1}^{n} I_0\left(\frac{r_j a_j}{\sigma_j^2} s\right) ds & \text{for } r_1 > 0, \ldots, r_n > 0, \\ 0 & \text{for other } r_1, \ldots, r_n, \end{cases} \qquad (5.12)$$

which depends on $2n+1$ parameters $a_1, a_2, \ldots, a_n, \sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2, q^2$. The form of this probability density, which is represented by an integral not expressible in terms of elementary functions (except for the cases $n = 1, 2$, which are of little interest), shows that calculations connected with a probability density of such a kind do not lend themselves to the usual numerical methods of solution. We observe that in the absence of signal the joint probability density of noise is given by expression (5.10), and in the presence of signal by expression (5.12), in which there occurs the integral of a combination of exponential and Bessel functions.

We observe that the probability distribution of amplitudes (5.12) arises also in problems connected with the passage of noise through multichannel systems, in which the frequency bands of the individual channels overlap each other thus causing stochastic connection between the noise elements.

Let us now consider a resultant signal with arbitrary correlation between the components of the voltage vectors. We shall assume, for the sake of simplicity, that the expectations of the components are equal to zero, and therefore the $2n$-dimensional probability density of the components has the form

$$p(X) = \frac{1}{(2\pi)^n \sqrt{D}} e^{-\frac{1}{2} X \mathfrak{M}^{-1} X}. \qquad (5.13)$$

In the problems of detection theory considered below the covariance matrix $\mathfrak{M} = \|m_{ij}\|$, where $m_{ij} = Mx_i x_j$ is the product-moment of the components x_i and x_j, is characterized by the following relations between the matrix elements:

$$\begin{aligned} m_{ij} &= m_{ji}, \\ m_{2k-1,2l-1} &= m_{2k,2l}, \\ m_{2k-1,2l} &= -m_{2l-1,2k}. \end{aligned} \qquad (5.14)$$

The first of these indicates that the covariance matrix $\mathfrak{M}$ is symmetrical. The second relation indicates that the two groups of components $x_1, x_3, \ldots, x_{2n-1}$ and $x_2, x_4, \ldots, x_{2n}$ are correlated in the same manner, and therefore the dispersions of the two components of the vector of the resultant voltage are equal

$$\sigma_k^2 = Mx_{2k-1}^2 = Mx_{2k}^2.$$

The last relation (5.14) indicates that these components are independent (for $k=l$)

$$Mx_{2k-1}x_{2k}=m_{2k-1,2k}=-m_{2k-1,2k}=-Mx_{2k-1}x_{2k},$$

i.e.,

$$m_{2k-1,2k}=0.$$

Let us find the n-dimensional probability density of the amplitudes of the signals. We observe, to this end, that the elements of the inverse matrix $\mathfrak{M}^{-1}=\|a_{ij}\|$ satisfy relations of the form (5.14). Therefore (5.13) can be rewritten in greater detail thus:

$$\begin{aligned} p(x_1,\dots x_{2n}) &= \\ \frac{1}{(2\pi)^n\sqrt{D}}\exp\Bigg\{-\frac{1}{2}\sum_{k,\,l=1}^{2n} & [a_{2k-1,2l-1}(x_{2k-1}x_{2l-1}+x_{2k}x_{2l}) \\ & +a_{2k-1,\,2l}(x_{2k-1}x_{2l}-x_{2l-1}x_{2k})]\Bigg\}. \end{aligned}$$

We pass in this expression to polar coordinates, by putting* $a_{2k-1,2l-1}=b_{kl}\cos\theta_{kl}$, $a_{2k-1,2l}=b_{kl}\sin\theta_{kl}$. Then the quadratic form in the exponent is found to be equal to

$$\begin{aligned} \Phi = \sum_{k,\,l=1}^{n} & [a_{2k-1,2l-1}r_kr_l\cos(\varphi_l-\varphi_k)+a_{2k-1,2l}r_kr_l \\ & \times\sin(\varphi_l-\varphi_k)]=\sum_{k=1}^{n}b_{kk}r_k^2+2\sum_{k<l}b_{kl}r_kr_l\cos(\varphi_l-\varphi_k-\theta_{kl}). \end{aligned}$$

By averaging the phase-and-amplitude probability density obtained

$$p(r_1,\varphi_1;\ r_2,\varphi_2;\ \dots;\ r_n,\varphi_n)=\frac{r_1\dots r_n}{(2\pi)^n\sqrt{D}}e^{-\frac{1}{2}\Phi(r_1,\dots,r_n;\ \varphi_1,\dots,\varphi_n)}$$

with respect to the phases, we arrive at an expression for the probability density of the amplitudes alone in terms of a multiple integral

---

*Thus we have

$$b_{kl}=b_{lk},\quad \theta_{kl}=-\theta_{lk}.$$

$$p(r_1, \ldots, r_n) = \frac{r_1 \ldots r_n}{\sqrt{D}} e^{-\frac{1}{2}\sum_{k=1}^{n} b_{kk} r_k^2}$$
$$\times \frac{1}{(2\pi)^n} \int_0^{2\pi} \ldots \int_0^{2\pi} \exp\left[-\sum_{k<l} b_{kl} r_k r_l \cos(\varphi_l - \varphi_k - \theta_{kl})\right]$$
$$\times d\varphi_1 \ldots d\varphi_n.$$

By writing such an integral as a series of repeated integrals and using the known formula from Bessel-function theory

$$e^{-z\cos\varphi} = \sum_{n=-\infty}^{\infty} (-1)^n I_n(z)\, e^{in\varphi},$$

where $I_n(z)$ is the modified Bessel function of the first kind and nth order, we obtain on reduction the following final expression for the required probability of the amplitudes where

$$p(r_1, \ldots, r_n) = \frac{1}{\sqrt{D}} r_1 \ldots r_n e^{-\frac{1}{2}\sum_{k=1}^{n} b_{kk} r_k^2}$$
$$\times \sum (-1)^{\sum_{k<l} \sum_{l=2}^{n} \nu_{lk}} \prod_{k=1}^{n-1} \prod_{l>k} \varepsilon_{\nu_{kl}} I_{\nu_{kl}} (b_{kl} r_k r_l)$$
$$\times \cos\left(\sum_{k=1}^{n-1} \sum_{l>k} \nu_{kl} \theta_{kl}\right), \qquad (5.15)$$

where

$$\varepsilon_{\nu_{kl}} = \begin{cases} 1 & \text{for} \quad \nu_{kl} = 0, \\ 2 & \text{for} \quad \nu_{kl} > 0. \end{cases}$$

The expression given emerges as a generalization of the probability density (5.12) for arbitrarily correlated signal and noise voltages. It is evident that in both cases the existence of explicit analytical expressions for the probability densities does not necessarily imply the feasibility of elementary calculations of the probability characteristics. It is just this fact that makes it necessary to have recourse in problems of detection theory to the Monte Carlo method.

In concluding this section we shall consider one more type of signal which occurs when at the receiver output there is a threshold-type device. The latter's action is as follows: any voltage exceeding the threshold level c gives rise to a standard unit signal while a voltage smaller than the level c is not recorded, i.e., gives a "zero" signal. Thus, as a result of the passage of signal through a threshold-type device the signal may be said to be "quantized," in that it assumes only two values, which can be termed the "one" and "zero" values. In this case the signal observed at the receiver output is a sequence of zeros and ones. The probabilistic description of a "quantized" signal is carried out by assigning either the probability of the occurrence of one or other of the two possible symbols if the successive elements of the resultant signal are independent, or the probabilities of combinations of zeros and ones in the case of correlation of the signal elements. In the first case we shall denote by p_l the probability that the amplitude of the lth component exceeds the threshold level c, and by $q_l = 1 - p_l$ the probability that this component falls below the threshold. Since the amplitudes of the voltages are distributed according to the Rayleigh law, the probabilities of interest to us can be written explicitly thus:

$$p_l = \mathrm{P}(r_l > c) = \int_c^\infty \frac{x}{\sigma_b^2} e^{-\frac{x^2}{2\sigma_l^2}} dx = e^{-\frac{c^2}{2\sigma_l^2}},$$

$$q_l = 1 - e^{-\frac{c^2}{2\sigma_l^2}}.$$

The probabilities of various combinations of zeros and ones are found without difficulty. The probability that an n-element signal consists of ones at the places $i_1, i_2, \ldots, i_\alpha$ and zeros at the remaining places is equal to

$$p_{i_1, \ldots, i_\alpha} = \prod_{k=1}^{\alpha} p_{i_k} \cdot \prod_{j \neq i_1, \ldots, i_\alpha} q_j.$$

A more complex picture is obtained in the case of correlated elements. In this case the description of the signal is accomplished by assigning the joint probability of occurrence of ones and zeros at the corresponding places, this probability being denoted by $p(i_1, i_2, \ldots, i_\alpha; j_1, \ldots, j_\beta)$ where $\alpha + \beta = n$. The evaluation of such a probability is carried out by integrating the probability density of the amplitudes over the corresponding region:

$$p(i_1, \ldots, i_\alpha; j_1, \ldots, j_\beta) = \int_c^\infty dr_{i_1} \ldots \int_c^\infty dr_{i_\alpha} \int_0^c dr_{j_1} \ldots \int_0^c dr_{j_\beta} p(r_1, \ldots, r_n).$$

2. FORMULATION OF THE BASIC PROBLEMS OF DETECTION THEORY

The problems formulated in this section arise in estimating the performance of various methods of extracting a useful signal from a noise background.

As a transmitted message passes through the communication channel and is amplified in the receiver, it is polluted by the channel noise and the receiver noise, and in such distorted form arrives at the recipient who assesses the message content on the basis of the voltage amplitudes. In many cases, for example in long-distance communication, the intensity of the signals proves to be so small that their very presence becomes uncertain, and the first problem the recipient has to solve consists in assessing whether useful signal is contained in the observed voltage or whether the latter is generated by noise alone. The random nature of the amplitudes of the voltages of which the incoming messages consist indicates that their complete characteristics will be contained in distribution laws. These laws will be different according to whether useful signal is present and at the output of the receiver there is a mixture of signal and noise, or whether there is only noise. It has been indicated in the previous section that in problems of detection theory this difference shows in the values of the parameters of the distributions. Thus the initial $(2n)$-dimensional normal distribution of the components of the vectors of the voltages is assigned by the covariance matrix $\mathfrak{M}$, which in the presence of signal is equal to the sum of the covariance matrices of signals $\mathfrak{M}_S$ and noise $\mathfrak{M}_N$ respectively. In the absence of signal this matrix is equal to $\mathfrak{M}_N$. Of course, the amplitude distributions obtained from these normal distributions also differ in the values of the parameters; by assuming that, in the corresponding probability density p_{SN} (of the mixture of signal and noise), the parameters generated by the elements of the matrix $\mathfrak{M}_S$ are equal to zero we arrive at the probability density p_N of the distribution of noise.

Thus the problem of detecting signals on a noise background reduces to testing, on the basis of observed values of the signal received, a hypothesis on the values of parameters of its distribution. Such problems are studied in mathematical statistics (see [12, 29, and 43]).

We shall consider two methods of solution of this problem. One of them has been developed by Neumann and Pearson, the other by Wald. The initial ideas of the two approaches are the same and consist in the following.

Let ξ be a random quantity which can be in general multidimensional $\xi=(\xi_1, \xi_2, \ldots, \xi_m)$ with probability density $p(X; \alpha)$, depending on the parameter α.* The functional form of the probability density is determined but the values of the parameters are unknown. As to the possible values of the parameters two hypotheses, H_0 and H_1, are permissible. These are of two kinds: simple or composite. They are termed simple when the hypothesis H_0 consists in assuming $\alpha=\alpha_0$ while the hypothesis H_1 consists in assuming $\alpha=\alpha_1$. The hypotheses are said to be composite when at least one of them allows the parameter α to assume several values. An example of the last situation is the following alternative: according to the hypothesis H_0 we have $\alpha=0$, and according to the hypothesis H_1 we have $\alpha>0$. We shall be concerned, below, with simple hypotheses.

Which of the two hypotheses is valid is assessed on the basis of results of drawing sample values of ξ, which has given probability density $p(X; \alpha)$. A sampling point will be denoted $X=(x_1, x_2, \ldots, x_m)$; the set of all such possible points belongs to an m-dimensional sampling space. The statistical methods considered for testing hypotheses consists in dividing the sampling space into subspaces (in Neumann's and Pearson's theory there are two of them, and in Wald's sequential analysis three), such that if the sampling point is found in one of them, the hypothesis H_0 is rejected (this is called the critical subspace), if it is found in another the hypothesis is accepted, and finally (in sequential analysis), if the sampling point is found in the third subspace, the experiment is continued owing to insufficient data.

The random nature of the sampling point may lead to the following types of wrong decisions as to the validity of the hypotheses: we may accept the hypothesis H_1 although it is H_0 that is valid (error of the first kind), or else we may reject the hypothesis H_1 in spite of the fact that it is valid (error of the second kind).

Let us illustrate what has been said in terms of detection theory: the recipient of the message, having measured the amplitudes of the voltages at the output of the receiver, must establish whether they are caused by the action of noise only (the hypothesis H_0) or whether they contain useful signal in addition to noise (the hypothesis H_1). Suppose the signal is actually present; as the recipient of the message has no previous knowledge of this, he may, not having detected the signal, mistakenly reject the correct hypothesis. On the other hand, suppose there is no signal; the recipient having "read" a nonexistent message may mistakenly take noise for signal. To have the probabilities of both errors small at the same time is impossible. It is natural therefore to fix the probability of mistaking noise for signal (i.e., the probability of error of the first kind) and to attempt to choose the critical region in such a manner that the probability of loss of signal (i.e., the probability of error of the second kind) be a minimum. The solution of this problem is dealt with in Neumann and Pearson's theory.

---

*To simplify the notation, we assume α to be a unidimensional quantity. The extension of the theory to the case of several parameters is evident.

Thus there are two alternative hypotheses: one of them, H_0, consists in assuming that sampling has been carried out from a general set with probability density $p(X;\ \alpha_0)$, i.e., that the parameter α, which uniquely defines the probability density, has the fixed value α_0; the other hypothesis, H_1, assumes that the parameter is equal to α_1. Let us fix the probability, p_{I}, of error of the first kind. We shall define the critical region as the set of points X that satisfy the inequality $p(X;\ \alpha_1) \geqslant cp(X;\ \alpha_0)$, where the constant c, which we shall term in the sequel threshold, is found from the condition that the probability p_{I} be equal to its assigned value. It may be shown that such a choice of the critical region is optimum in the sense that the probability of error of the second kind will be a minimum.

Let us introduce the notation

$$l(X) = \frac{p(X;\ \alpha_1)}{p(X;\ \alpha_0)},$$

which in statistics is called the likelihood ratio. Then the optimum criterion for deciding between the two hypotheses (this is the content of the basic theorem of Neumann and Pearson's theory) will be expressed thus: if

$$l(X) \geqslant c,$$

the hypothesis H_1 is accepted; in the opposite case the hypothesis H_0 is accepted.

The probabilities of errors arising as a result of adopting this criterion are expressed by integrals

$$p_{\text{I}} = \int\limits_{l(X) \geqslant c} p(X;\ \alpha_0)\,dX, \quad p_{\text{II}} = \int\limits_{l(X) < c} p(X;\ \alpha_1)\,dX, \tag{5.17}$$

in which the domain of integration is the subset of the corresponding sampling space defined by one of the inequalities $l(X) \geqslant c$ or $l(X) < c$. The order of multiplicity of the integrals is equal, of course, to the number of dimensions of the random quantity X.

We shall consider the application of Neumann and Pearson's theory to the problem of detecting a signal on a noise background. The recipient of the message has available n voltage amplitudes on the basis of which, by knowing the statistical characteristics of signals and noise, he has to establish whether useful signals are present or absent. The best method of doing it consists, according to Neumann and Pearson, in substituting the observed values of the amplitudes in the likelihood ratio and comparing its value with the threshold c. This means that the optimum receiver must comprise a device to accomplish on the input voltages the operations indicated by the functional form of the likelihood ratio. As a result of using this algorithm, and for a given probability of error of the first kind (which in technical applications is called the probability of false alarm), we reduce to a minimum the probability of error of the second kind (i.e., the probability of mistaking signal for noise). We usually reckon not in terms of the latter probability but in terms of the probability of the opposite event: the detection of signal. The probability of false alarm and the probability of detection will be denoted F and D respectively. They are expressed, in terms of the probability densities of noise $P_N(X)$ and of signal and noise combined $P_{SN}(X)$, in the following manner:

$$F = \int\limits_{l(X) \geqslant c} P_N(X)\,dX, \quad D = \int\limits_{l(X) \geqslant c} P_{SN}(X)\,dX. \tag{5.18}$$

The procedure described gives for a given probability F the highest value of the probability D.

An assessment of optimum methods of extracting signal from noise requires the evaluation of these probabilities; this need is even greater when, owing to the complexity of the optimum method, one has recourse instead of it to different methods. The difficulties of evaluating F and D arise mainly from the complex structure of the region of integration, which is defined as the set of points for which the likelihood ratio $l(X)$ is larger than the threshold c, as well as from the high order of multiplicity of the integrals.

We have considered above a number of typical forms of probability densities of the voltage amplitudes. We shall now write down the corresponding likelihood ratios, which will enable us to illustrate typical features of the integrals expressing the probabilities F and D. We shall firstly isolate the parameter of the probability density the values of which distinguish the various hypotheses from each other. In the case of random signal we shall take for such a parameter the quantity q, namely the ratio of the standard deviations of signal and noise, σ_S and σ_N, (briefly the signal-to-noise ratio); if however the signal is not random, the parameter will be s, namely, the ratio of the signal amplitude to the standard deviation of noise. Then the hypothesis H_0 consists in assuming that there is no signal $q=0$ (or $s=0$), and that the voltage observed is caused by noise alone. According to the alternative hypothesis there is signal, which is characterized by a fixed value of the parameter: q_1 for a random signal and s_1 for a nonrandom one. Actually the value of the parameter is, as a rule, unknown to the message recipient, and therefore, strictly speaking, we have to test the simple hypothesis $H_0\{q=0\}$ against the composite one $H_1\{q>0\}$. The corresponding theory, however, has been little developed and is inapplicable to many types of signals. The following argument provides justification for using, in such situations, the theory of simple-hypothesis testing. Let the minimum acceptable detection probability D_1 be prescribed on the basis of various relevant considerations (its value varies from 0.1 to 0.99 in various applications). We shall term threshold signal the value of the signal-to-noise ratio that ensures this probability and shall denote it by q_1 (or, for a nonrandom signal, by s_1). If now the receiver is so designed that it best realizes this detection probability, the corresponding threshold signal will be a minimum.* For large values of the signal-to-noise ratio (and only these are of interest here) the incoming signal is more intense and the minimum acceptable detection probability D_1 is ensured although the receiver does not process such signals optimally. The above considerations are of an intuitive character and, of course, do not provide a rigorous proof of the possibility of using the theory of simple-hypothesis testing. However, as experiments show, they are adequate for certain engineering applications.

The determination of the threshold signals is not the only computing problem of detection theory. Of no less importance is the plotting of "detection curves," i.e., of the detection probability D as a function of the signal-to-noise ratio. To do this, having chosen a receiver that best ensures the detection of signal with false-alarm probability equal to an assigned value D_1, and having thus fixed the parameter q_1 (or s_1), one considers the detection of signals the probability densities of which are characterized by different values of the parameter q. In a number of elementary cases such functions $D=D(q;\ F)$ can be obtained explicitly. Many problems, however, require a large amount of computation for obtaining numerical solutions. Examples of such problems will be repeatedly met in the sequel.

---

*The minimization of the threshold signal is important in engineering, since it enables us to reduce the energy expended in the transmitter for exciting the oscillations carrying the useful signal.

Let us explicitly introduce the parameter q (or s for nonrandom signals) in the notation for the probability density of signal and noise combined, so that

$$P_{SN}(X) = p(X;\ q);$$

then the probability density of noise alone will be

$$P_N(X) = p(X;\ 0).$$

The likelihood ratio will also depend on the signal-to-noise ratio:

$$l(X;\ q) = \frac{p(X;\ q)}{p(X;\ 0)}.$$

We shall begin our enumeration of likelihood ratios with the case of noiselike signals described by the probability density (5.10). By taking the noise variance as unity, we express the variance σ_k^2 in the form

$$\sigma_k^2 = 1 + a_k^2 q^2, \tag{5.19}$$

where a_k is the "envelope" of the sequence of amplitudes of the useful signal. Then the probability density of the amplitudes of signal and noise combined is equal to

$$P(r_1, \ldots, r_n;\ q) = \prod_{k=1}^{n} \frac{r_k}{1 + a_k^2 q^2} e^{-\frac{r_k^2}{2\left(1+a_k^2 q^2\right)}},$$

and the probability density of noise alone is equal to

$$P(r_1, \ldots, r_n;\ 0) = \prod_{k=1}^{n} r_k e^{-\frac{r_k^2}{2}}.$$

Therefore the likelihood ratio is equal to

$$l(r_1, \ldots, r_n;\ q) = \frac{1}{\prod_{k=1}^{n}\left(1 + a_k^2 q^2\right)} e^{\frac{q^2}{2} \sum_{k=1}^{n} \frac{a_k^2}{1+a_k^2 q^2} r_k^2}. \tag{5.20}$$

The hypothesis of existence of signal is accepted if the likelihood ratio satisfies the inequality

$$l(r_1, \ldots, r_n;\ q) \geqslant c.$$

By passing to logarithms and carrying out elementary transformations, we arrive at the equivalent inequality

$$\sum_{k=1}^{n} \frac{a_k^2}{1+a_k^2q^2} r_k^2 \geqslant \frac{2}{q^2} c + \frac{2}{q^2} \sum_{k=1}^{n} \ln\left(1+a_k^2q^2\right). \tag{5.21}$$

The right-hand side of this inequality will be denoted, for brevity, c_q.

Thus the detection of noiselike signal is best accomplished by means of a (weighted) sum of the squares of the amplitudes. Such processing of signals is called square-law processing. The probabilities F and D are expressed by improper n-tuple integrals taken over the region external to the ellipsoid

$$\text{Э}_q = \left\{ r_1, \ldots, r_n;\ \sum_{k=1}^{n} \frac{a_k^2}{1+a_k^2q^2} r_k^2 \geqslant c_q \right\},$$

i.e.,

$$F = \int_{\text{Э}_q} \cdots \int \prod_{k=1}^{n} r_k e^{-\frac{r_k^2}{2}} \, dr_1 \ldots dr_n;$$

$$D = \int_{\text{Э}_q} \cdots \int \prod_{k=1}^{n} \frac{r_k}{1+a_k^2q^2} e^{-\frac{r_k^2}{2\left(1+a_k^2q^2\right)}} \, dr_1 \ldots dr_n.$$

The threshold signal q_1 and the corresponding threshold level c ensuring the assigned false-alarm probability must be found by solving the system of equations

$$F(q_1;\ c) = F, \quad D(q_1;\ c) = D_1. \tag{5.22}$$

In order to determine the detection probability D as a function of the signal-to-noise ratio q we need to evaluate the integral

$$\mathcal{D}(q;\ q_1;\ c) = \int_{\text{Э}_{q_1}} \cdots \int \prod_{k=1}^{n} \frac{r_k}{1+a_k^2q^2} e^{-\frac{r_k^2}{2\left(1+a_k^2q^2\right)}} \, dr_1 \ldots dr_n.$$

We observe that

$$\mathcal{D}(0;\ q_1;\ c)=F;\quad \mathcal{D}(q_1;\ q_1;\ c)=D_1.$$

It can easily be verified that the above integrals can be expressed in terms of tabulated functions, as combinations of χ^2-distributions with an even number of degrees of freedom. If the signal envelope is constant (i.e., $a_k=a_l$ for all k and l), the probabilities D and F are expressed by the χ^2-distribution with $2n$ degrees of freedom.

Therefore, the use of the Monte Carlo method for evaluating the characteristics of square-law processing is not advisable. However, owing to the fact that a simple analytical representation of the required probabilities is available, this problem can be used to assess the power of the Monte Carlo method: to verify the quality of pseudo-random numbers, to estimate the rate of convergence, etc.

Let us now consider the detection of signals that differ from one signal element to another by a constant factor a_k, the noise voltages in different elements being uncorrelated with each other. The probability density of the amplitudes of signal and noise combined is equal, for an n-element message, to

$$P(r_1,\ \ldots,\ r_n;\ q)=$$
$$\prod_{k=1}^{n}\frac{r_k}{\sigma_k^2}e^{-\frac{r_k^2}{2\sigma_k^2}}\int_0^\infty\frac{s}{q^2}e^{-\left(\frac{1}{q^2}+\sum\frac{a_k^2}{\sigma_k^2}\right)\frac{s^2}{2}}\prod_{j=1}^{n}I_0\left(\frac{r_ja_j}{\sigma_j^2}s\right)ds,$$

while the probability density of the amplitudes of the noise voltages alone is equal to

$$P(r_1,\ \ldots,\ r_n;\ 0)=\prod_{k=1}^{n}\frac{r_k}{\sigma_k^2}e^{-\frac{r_k^2}{2\sigma_k^2}}. \tag{5.23}$$

The likelihood ratio is expressed by the integral

$$l(r_1,\ \ldots,\ r_n;\ q)=$$
$$\int_0^\infty\frac{s}{q^2}e^{-\left(\frac{1}{q^2}+\sum\frac{a_k^2}{\sigma_k^2}\right)\frac{s^2}{2}}\prod_{j=1}^{n}I_0\left(\frac{r_ja_j}{\sigma_j^2}s\right)ds. \tag{5.24}$$

The optimum receiver, according to (5.24), evaluates for each received pulse the functions $I_0\left(\frac{r_j a_j}{\sigma_j^2} s\right)$, $j = 1, \dots, n$ and then averages their product with weight $\frac{s}{q^2}\exp\left\{-\left(\frac{1}{q^2}+\sum \frac{a_k^2}{\sigma_k^2}\right)\frac{s^2}{2}\right\}$. The errors arising in such a method of reception are characterized by the probabilities F and D, equal respectively to

$$F = \int_{l(r_1, \dots, r_n;\, q) > c} \dots \int \prod_{k=1}^{n} \frac{r_k}{\sigma_k^2} e^{-\frac{r_k^2}{2\sigma_k^2}} dr_1 \dots dr_n,$$

$$D = \int_{l(r_1, \dots, r_n;\, q) > c} \dots \int \prod_{k=1}^{n} \frac{r_k}{\sigma_k^2} e^{-\frac{r_k^2}{2\sigma_k^2}} dr_1 \dots dr_n$$

$$\times \int_0^\infty \frac{s}{q^2} e^{-\left(\frac{1}{q^2}+\sum_{k=1}^{n}\frac{a_k^2}{\sigma_k^2}\right)\frac{s^2}{2}} \prod_{j=1}^{n} I_0\left(\frac{r_j a_j}{\sigma_j^2} s\right) ds.$$

Just as in the case of square-law processing, the threshold signal q_1 and the threshold level c are determined by the system of equations (5.22). The detection probability D as a function of q is found by evaluating the integral

$$\mathscr{D}(q; q_1; c) = \int_{l(r_1, \dots, r_n;\, q_1) > c} \dots \int P(r_1, \dots, r_n;\, q)\, dr_1 \dots dr_n. \qquad (5.25)$$

In this case, it is no longer possible to reduce the calculations needed to elementary or tabulated functions, while the use of quadrature formulas is made difficult by the high order of multiplicity of the integral. Thus the Monte Carlo method emerges here as the only acceptable method of computation.

Let us consider the form of the likelihood ratio for arbitrary correlation between the components of the signal and noise voltages, when the successive signals in the message are connected not functionally but stochastically. The corresponding joint probability density of the amplitudes has been obtained in Section 1, the formula (5.15). We shall qualify with an index 1 quantities corresponding to signal and noise combined and with an index 0 quantities corresponding to noise alone. Then the likelihood ratio is found to be equal to

$$l(r_1, \ldots, r_n;\ q) = \sqrt{\frac{D_0}{D_1}}\, e^{-\frac{1}{2}\sum\left(b^1_{kk}-b^0_{kk}\right)r_k^2}$$
$$\times \frac{\sum(-1)^{\sum\sum\nu_{kl}}\prod\limits_{k=1}^{n-1}\prod\limits_{l>k}\varepsilon_{\nu_{kl}}I_{\nu_{kl}}\left(b^1_{kl}r_kr_l\right)\cos\left(\sum\sum\nu_{kl}\theta^1_{kl}\right)}{\sum(-1)^{\sum\sum\nu_{kl}}\prod\limits_{k=1}^{n}\prod\limits_{l>k}\varepsilon_{\nu_{kl}}I_{\nu_{kl}}\left(b^0_{kl}r_kr_l\right)\cos\left(\sum\sum\nu_{kl}\theta^0_{kl}\right)}.$$

In the particular case when the noise voltages are independent, this expression is somewhat simplified

$$l(r_1, \ldots, r_n;\ q) = \frac{1}{\sqrt{D}}\, e^{-\frac{1}{2}\sum(b_{kk}-1)r_k^2} \sum(-1)^{\sum\sum\nu_{kl}}$$
$$\times \prod_{k=1}^{n-1}\prod_{l>k}\varepsilon_{\nu_{kl}}I_{\nu_{kl}}(b_{kl}r_kr_l)\cos\left(\sum\sum\nu_{kl}\theta_{kl}\right).$$

The false-alarm probability F and the detection probability D are expressed by means of the corresponding probability densities in the usual manner.

Let us pass to the other method of signal detection in noise. This is based on sequential analysis, a development of mathematical statistics, and has attracted much attention in recent years; a great number of papers has been devoted to its application to the group of problems of interest to us. In sequential analysis, in contrast to Neumann and Pearson's theory, the amount of sampling used to test a statistical hypothesis is not fixed in advance, but is determined as a result of the test carried out. The advantage of this method is due to the reduced number of tests needed, which leads to economy in the time spent on the experiment.

The practical realization of the method of sequential analysis in detection problems consists in the following: the hypothesis H_0 is that signal is absent, and therefore the signal-to-noise ratio, i.e., the parameter on which the probability density of the voltage amplitudes depends, is equal to zero, and the alternative hypothesis H_1, that the voltages observed are caused by the combined action of noise and signal, the latter being characterized by a fixed value of the signal-to-noise ratio. Let us denote by

$$P_1(r_1;\ q),\ P_2(r_1,\ r_2;\ q),\ \ldots,\ P_k(r_1,\ \ldots,\ r_k;\ q),\ \ldots$$

and

$$P_1(r_1;\ 0),\ P_2(r_1,\ r_2;\ 0),\ \ldots,\ P_k(r_1,\ \ldots,\ r_k;\ 0),\ \ldots$$

the k-dimensional joint probability densities of signal and noise combined and of noise alone. The likelihood ratio will be denoted by

$$l_k(r_1, \ldots, r_k; q) = \frac{P_k(r_1, \ldots, r_k; q)}{P_k(r_1, \ldots, r_k; 0)}.$$

Let us assign the positive numbers A and $B\,(A > B)$. The tests to establish which of the two hypotheses H_0 and H_1 is valid will be carried out in the following manner. As a first step, having measured the amplitude r_1, we evaluate $l_1(r_1; q)$.

If it is found that $l_1(r_1; q) > A$, we accept the hypothesis H_1. If the inequality $B > l_1(r_1; q)$ is satisfied we accept the hypothesis H_0. Finally, if $B < l_1(r_1; q) < A$, the decision as to the validity of one or other hypothesis is postponed, and further tests are carried out. This means that the next amplitude r_2 is measured and $l_2(r_1, r_2, q)$ is evaluated. Again there are three possibilities

$$l_2(r_1, r_2; q) > A, \quad l_2(r_1, r_2; q) < B \text{ and } B < l_2(r_1, r_2; q) < A.$$

According to which of these inequalities is satisfied, we proceed in a similar manner to the above. If at the $(k-1)$th sequential test no decision has yet been reached, i.e., if

$$B < l_n(r_1, \ldots, r_n; q) < A, \qquad n = 1, 2, \ldots, k-1,$$

then, at the kth step, we add to the amplitudes $r_1, r_2, \ldots, r_{k-1}$ the amplitude r_k and evaluate $l_k(r_1, \ldots, r_k; q)$ which is again compared with both thresholds A and B. The thresholds A and B are so chosen that the probabilities of errors of the first and second kind (i.e., the probability of false alarm and the probability of signal rejection) are equal to fixed values; in fact

$$A = \frac{D}{F}, \qquad B = \frac{1-D}{1-F}.$$

It can be shown that for many important types of distribution the probability of the process of sequential testing ending with the acceptance of one of the two hypotheses is equal to unity. This is true, in particular, of independent observations, when the k-dimensional joint probability density is expressed as the product of the probability densities of the individual arguments:

$$P_k(r_1, \ldots, r_k; q) = P_1(r_1; q)\,P_1(r_2; q) \ldots P_1(r_k; q).$$

In connection with the fact that sequential analysis is used for reducing the amount of sampling (or else, which is the same, the observation time), the main characteristic of performance of the method is the probability distribution of the serial number of the end step of the process, i.e., the distribution of the serial number of the step at which, as a result of analysis of the amplitudes observed, one of the two concurrent hypotheses is accepted. It suffices in certain cases to restrict ourselves to the first two moments of this distribution.

There are four possibilities of ending the sequential analysis at the νth step: the correct hypothesis H_0 is accepted or is rejected, the correct hypothesis H_1 is accepted or is rejected. Let us denote the corresponding probabilities by the symbols

$$p_0(\nu;\, 0), \qquad p_0(\nu;\, q)$$

for the first two possibilities and

$$p_q(\nu;\, q), \qquad p_q(\nu;\, 0)$$

for the last two possibilities, and let us recall that according to the hypothesis H_1 a signal of intensity q is present in the voltage observed.

Let us consider in greater detail the events the probabilities of which are described as $p_0(\nu;\, 0)$ and $p_0(\nu;\, q)$. The occurrence of any of these means that the sequence of amplitudes $r_1, \ldots, r_\nu$ with joint probability density $P_k(r_1, \ldots, r_k;\, q)$, $k = 1, \ldots, \nu$, is such that the following inequalities are satisfied: either

$$B < l_k(r_1, \ldots, r_k;\, q) < A, \qquad k = 1, \ldots, \nu - 1$$

and

$$l_\nu(r_1, \ldots, r_\nu;\, q) < B$$

(when the first of the events is realized), or

$$\begin{gathered} B < l_k(r_1, \ldots, r_k;\, q) < A, \qquad k = 1, \ldots, \nu - 1, \\ l_\nu(r_1, \ldots, r_\nu;\, q) > A \end{gathered}$$

(when the second event is realized). The meaning of the events having probabilities $p_q(\nu;\, 0)$ and $p_q(\nu;\, q)$ may be illustrated in a similar manner.

In order to clarify the typical difficulties that are met in evaluating the above probabilities, we shall consider the simple version when the amplitudes $r_1, r_2, \ldots, r_n, \ldots$ are independent and the likelihood ratio is expressed in the form

$$l_n(r_1, \ldots, r_n;\, q) = \prod_{k=1}^{n} l_1(r_k;\, q).$$

It is convenient, in this case, to pass to logarithms, as a result of which the process of sequential detection reduces to adding the random quantities $\zeta_k = \ln l_1(r_k;\, q)$. If the following inequalities

$$\ln B < \zeta_1 + \zeta_2 + \ldots + \zeta_n < \ln A, \qquad k = 1, \ldots, \nu - 1,$$

are satisfied at the first ν-1 steps, and one of the inequalities

$$\begin{gathered} \zeta_1 + \ldots + \zeta_\nu > \ln A = a, \\ \zeta_1 + \ldots + \zeta_\nu < \ln B = b, \end{gathered}$$

is satisfied at the ν-th step, the process ends with one of the two hypotheses being accepted. Owing to the statistical dependence between the sequential sums $\zeta_1 + \ldots + \zeta_n$ and $\zeta_1 + \ldots + \zeta_m$, $m > n = 1, 2, \ldots$, the calculation of the probabilities of interest to us reduces to evaluating n-tuple integrals.

We shall write down, for example, the formulas for the probabilities of the serial number of the final step of sequential analysis for a noiselike signal. To make matters simpler, we restrict ourselves to the case of a constant signal envelope

$$\ln l_k(r_1, \ldots, r_k;\, q) = \frac{q^2}{2(1+q^2)} \sum_{i=1}^{k} r_i^2 - k \cdot \ln(1+q^2).$$

The upper and lower thresholds of the quantity $\sum \ln l_k$ are expressed in terms of A and B and, after elementary transformations, are found to depend on the serial number of the test

$$\begin{aligned} a_k &= 2\frac{1+q^2}{q^2}[\ln A + k \cdot \ln(1+q^2)], \\ b_k &= 2\frac{1+q^2}{q^2}[\ln B + k \cdot \ln(1+q^2)], \end{aligned}$$

so that the process of analysis is continued until the inequalities

$$b_k < \sum_{i=1}^{k} r_i^2 < a_k$$

are satisfied, and is ended with accepting the corresponding hypothesis, when at the νth step the quantity $\ln l_\nu(r_1, r_2, \ldots, r_\nu;\, q)$ lies outside the limits of the interval (b_ν, a_ν). The quantities r_i^2 are distributed according to the exponential law

$$p(x)=\begin{cases} \frac{1}{d}e^{-\frac{x}{d}}, & x \geqslant 0, \\ 0\ , & x < 0, \end{cases}$$

where $d=1$ if only noise is present, and $d=1+q^2$ if there is signal and noise combined. From this we obtain an expression for the probability of ending the process at the νth step by accepting the hypothesis H_1 of presence of signal, $p_0(\nu;\,q)$ or $p_q(\nu;\,q)$:

$$\left.\begin{matrix} p_0(\nu;\,q) \\ p_q(\nu;\,q) \end{matrix}\right\}= \int\limits_{b_\nu-(x_1+\ldots+x_{\nu-1})}^{\infty} P(x_\nu)\,dx_\nu \int\limits_{b_{\nu-1}-(x_1+\ldots+x_{\nu-2})}^{a_{\nu-1}-(x_1+\ldots+x_{\nu-2})} P(x_{\nu-1})\,dx_{\nu-1}\ \ldots$$
$$\ldots \int\limits_{b_3-x_1-x_2}^{a_3-x_1-x_2} P(x_3)\,dx_3 \int\limits_{b_2-x_1}^{a_2-x_1} P(x_2)\,dx_2 \int\limits_{b_1}^{a_1} P(x_1)\,dx_1. \tag{5.26}$$

In a similar manner we can express the probability of accepting the hypothesis H_0 of absence of signal. It is evident that the evaluation of such integrals does not involve any difficulty of principle, but it becomes extremely laborious as the number n increases. For other distributions such calculations may prove impossible to perform. We shall consider, for example, the investigation of the probability distribution of the quantity ν for a deterministic signal, described (when combined with noise) by the probability density (5.11). In this case, the form of the integrals (5.26) becomes substantially more complicated. In fact, the likelihood ratio at the first step has the form

$$l(r;\,s)=e^{-\frac{s^2}{2}}I_0(rs),$$

where the amplitude r is distributed according to Rice's law

$$p_s(x)=xe^{-\frac{x^2+s^2}{2}}I_0(rs),$$

if a signal has been transmitted, or according to Rayleigh's law

$$p(x)=xe^{-\frac{x^2}{2}},$$

if noise alone is present.

Tests are continued until the sum

$$\ln I_0(r_1 s) + \ln I_0(r_2 s) + \ldots + \ln I_0(r_k s)$$

lies beyond the limits of the interval (b_k, a_k) where

$$a_k = \ln A + k\frac{s^2}{2},$$
$$b_k = \ln B + k\frac{s^2}{2}.$$

The probability density of the quantity $\zeta = \ln I_0(rs)$, where r is distributed according to either the Rayleigh or the Rice law, can be obtained in principle on the basis of known rules for evaluating the distribution function of a random quantity. Clearly, however, we cannot expect to be able to write down explicit analytical expressions for these probability densities. Therefore the evaluation in closed form of the distribution of ν is impossible and only the Monte Carlo method enables us to obtain numerical results.

We can calculate in a similar manner the characteristics of sequential detection of other types of signals.

3. PROCEDURE FOR THE SOLUTION OF THE MAIN PROBLEMS OF DETECTION THEORY

It has been shown in Section 2 that the solution of the main problems of detection theory reduces to evaluating multiple integrals.

We shall consider, for example, the method of calculating the detection characteristics of a noiselike signal in the case of square-law processing. The probability of false alarm F and the detection probability D are expressed in the following manner:

$$F = \mathscr{D}(0;\ q_1;\ c),$$
$$D = \mathscr{D}(q;\ q_1;\ c), \qquad q > 0,$$

where

$$\mathscr{D}(q;\ q_1;\ c) = \int \underset{\text{Э}_{q_1}}{\cdots} \int \prod_{k=1}^{n} \frac{r_k}{1 + a_k^2 q^2} e^{-\frac{r_k^2}{2\left(1 + a_k^2 q^2\right)}} \, dr_k,$$

the region of integration is

$$Э_{q_1}=\left\{r_1, \ldots, r_n;\ \sum_{k=1}^{n}\frac{a_k^2}{1+a_k^2q^2}r_k^2 \geqslant c_{q_1}\right\},$$

and the limiting threshold c_{q_1} is equal to

$$c_{q_1}=\frac{2}{q_1}\left[c+\sum_{k=1}^{n}\ln\left(1+a_k^2q^2\right)\right].$$

The evaluation of such integrals by the Monte Carlo method is accomplished by simulating a sequence of series of Rayleigh's pseudo-random numbers

$$\begin{array}{l} r_{11},\ r_{12},\ \ldots,\ r_{1n}, \\ r_{21},\ r_{22},\ \ldots,\ r_{2n}, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ r_{N1},\ r_{N2},\ \ldots,\ r_{Nn}, \end{array}$$

either with parameters $1+a_k^2q^2$ (for signal and noise combined) or with parameter 1 (for noise alone). From the elements of each row, which simulate the amplitudes received, we form the sums

$$\zeta_l=\sum_{k=1}^{n}\frac{a_k^2}{1+a_k^2q^2}r_{lk}^2,$$

which are compared with the threshold c_{q_1}. If the inequality

$$\zeta_l \geqslant c_{q_1}$$

is satisfied, unity is added to the sum ν in the counter of successful tests. After repeating this procedure N times, the required integrals are identified with the quantity

$$\frac{\nu}{N}.$$

The method can be simplified if we recall that the square of a Rayleigh-distributed quantity has an exponential distribution. Therefore, by generating pseudo-random numbers with this distribution, we avoid the operations of root-extraction and squaring to be carried out on each "Rayleigh" number. What has been said concerns the method of evaluating a single integral denoting either F or D. In reality, one has to find many such integrals, since in the majority of problems one has to find a threshold c such that the probabilities F and D are equal to their assigned values. Let us first consider the simpler case where the expected signal q_1 and the probability of false alarm F are fixed. To this end, an electronic computer is used, as is indicated above, to generate the random numbers

$$\zeta=\sum_{k=1}^{n}\frac{a_k^2}{1+a_k^2q^2}r_k^2,$$

where the r_k^2 are distributed according to an exponential law with density

$$p(x)=\frac{1}{2}e^{-\frac{x}{2}}$$

and represent the squares of the noise amplitudes. Each of the numbers ζ is entered in a histogram and after a sufficiently large number of tests, when the form of the empirical probability density has been determined with acceptable accuracy, one determines the threshold c_{q_1} (and with it also c) which is such that the probability that the quantity ζ exceeds it is equal to the assigned probability of false alarm F. After this, it remains to evaluate the probability of detection $\mathscr{D}(q;\ q_1;\ c)$ in the region of values of the parameter q which have been considered, and in particular for $q=q_1$. By virtue of the fact that the threshold c_{q_1} is determined from a histogram and not by directly solving the equation

$$\int_{c_{q_1}}^{\infty} p_\zeta(x)\,dx=F,$$

the expenditure of machine time is reduced.

Let us consider now the method of calculation of the threshold signals. The probabilities F and D are assigned:

$$F=\mathscr{D}(0;\ q_1;\ c),\quad D=\mathscr{D}(q_1;\ q_1;\ c).$$

and it is required to find the corresponding values of q_1 and c and then to calculate $\mathscr{D}(q;\ q_1;\ c)$. In comparison with the previous one this method presents considerable difficulty, owing to the fact that the likelihood ratio $l(r,\ q_1)$, characterizing the region of integration in both integrals (for F and D), depends on the unknown q_1. Therefore the solution of the system of equations determining q_1

and c is accomplished by a trial method: one assigns a value of q_1 and makes up a histogram of the quantity ζ (simulated on the assumption of the presence of noise alone), from which the threshold level c_{q_1} (and with it also c) is chosen. Then $\mathscr{D}(q_1;\ q_1;\ c)$ is evaluated; if the result is found to be larger (smaller) than the assigned probability D, q_1 is increased (decreased). Of course, this process is very laborious and a considerable amount of machine time is required to achieve satisfactory accuracy in the determination of q_1 and c. In particular cases, for example for a noiselike signal with constant "envelope" ($a_k = a_l$ for all k and l), the likelihood ratio does not depend on q_1 and then, having determined the limiting threshold, we find the threshold signal-to-noise ratio by solving the equation $\mathscr{D}(q_1;\ q_1;\ c) = D$.

We have considered above the calculation of the threshold signals and of the detection characteristics for the case of square-law processing of noiselike signals. The main features of the application of the Monte Carlo method are retained also with other types of signals, the only difference consisting in the methods of simulating the corresponding random quantities. These methods are considered in Chapter VII, in which Section 6 is devoted to multidimensional distributions met in detection theory. If, for example, the noise and signal components are correlated, then the calculation starts by constructing the matrix A which transforms independent normal quantities into correlated quantities with assigned covariance matrix $\mathfrak{M}$. The successive independent normal quantities, transformed by this matrix, are used to find the amplitudes; these are substituted in the likelihood ratio which is compared with the threshold, etc.

Finally, let us consider how the characteristics in sequential analysis are calculated. As we have seen in Section 3, the object of the calculation is to find the probability distribution of the number of steps needed to end the process. This is achieved by simulating the sequence of random quantities

$$r_1,\ r_2,\ \ldots,\ r_n,\ \ldots$$

with joint probability densities

$$P_1(r_1),\ P_2(r_1\ r_2),\ \ldots,\ P_n(r_1,\ \ldots,\ r_n),\ \ldots$$

of one or two types: either concerning signal and noise combined, or noise alone, and by calculating the frequencies of the events consisting in the following inequalities being satisfied

$$\begin{gathered} B < l_k(r_1,\ \ldots,\ r_k) < A, \qquad k = 1,\ \ldots,\ \nu - 1, \\ l_\nu(r_1,\ \ldots,\ r_\nu) > A, \end{gathered}$$

or

$$l_\nu(r_1, \ldots, r_\nu) < B.$$

This means that at each step of the sequential analysis the calculated value of the likelihood ratio is compared with thresholds A and B, and when it lies beyond the limits of the interval (B, A) the serial number of this step is entered in the histogram. Then the process is repeated from the start. In the case of independent amplitudes, the probability is unity that the analysis ends after a finite number of steps (a similar statement would not be true for correlated amplitudes). It may be found that the working memory of the machine is not sufficient for the histogram. To avoid this difficulty, one can assign a serial number ν_0 and assume that all tests that continue further than the $(\nu_0 - 1)$th step have ended at the ν_0th step. Then the calculated probability p_{ν_0} is the sum of the probabilities of ending the process at the ν_0th, $(\nu_0 + 1)$th, ... steps.

4. OTHER PROBLEMS

In information-transmission theory the Monte Carlo method can be used to solve a wide group of problems in addition to those considered above. The most typical of them will be discussed below.

Detection of a Target and Determination of Its Position by Means of Counters of Quantized Signals [48b]

Let us consider the problem of radar detection of a target on an azimuth scan (on a sector at a fixed distance from a search radar). In a region not containing the target only Rayleigh-distributed noise pulses are present. The amplitudes of signals reflected from the target are assumed to be constant (the signal "does not fluctuate") and therefore the amplitudes of the resultant of signal and noise will be independent and distributed according to the Rice law.

Suppose the receiver accomplishes quantization, i.e., its output is a unit signal if the input voltage exceeds a level x_0 and a zero signal if the input voltage does not exceed x_0. As a result an azimuth scan is found to be filled with a sequence of zeros and ones.

Let the probability of occurrence of a noise pulse be denoted by p_N. By taking the variance of noise as the measurement unit, we have

$$p_N = \int_{x_0}^{\infty} x e^{-\frac{x^2}{2}} dx = e^{-\frac{x_0^2}{2}}. \tag{5.27}$$

The probability of occurrence of a signal pulse, p_S, depends on the signal amplitude and is equal to

$$p_S(s) = \int_{x_0}^{\infty} x e^{-\frac{x^2+s^2}{2}} I_0(sx)\,dx. \tag{5.28}$$

Since $p_S > p_N$, the occurrence of target will result in an increased number of ones in the corresponding region of the azimuth scan. Several criteria based on this consideration may be suggested. We shall indicate two of them.

a) Series type criterion. A counter of pulses is placed at the receiver output. As long as there are only zeros on the scan the counter remains set at zero. When the first one occurs on the scan, a unit is sent to the counter. Additional rules will be formulated by denoting by μ_k the outcome of a binomial test at the kth position of the scan, and by l_k the value registered by the counter corresponding to the kth position. If one of the conditions

$$\left.\begin{array}{ll} 1.\ 0 \leqslant l_n \leqslant 3 & \text{and} \left\{\begin{array}{ll} \text{either} & \mu_{n-3}\mu_{n-2}\mu_{n-1}\mu_n = 0100, \\ \text{or} & \mu_{n-2}\mu_{n-1}\mu_n = 000, \end{array}\right. \\ 2.\ 3 < l_n \leqslant 6 & \text{and } \mu_{n-2}\mu_{n-1}\mu_n = 000, \\ 3.\ 6 < l_n & \text{and } \mu_{n-3}\mu_{n-2}\mu_{n-1}\mu_n = 0000, \end{array}\right\} \tag{5.29}$$

is satisfied, the counter is returned to zero ($l_{n+1} = 0$), i.e., the end of the region of signals reflected from the target is recorded. For the remaining combinations of zeros and ones on the scan and values of the counter, one unit is added to latter, i.e., $l_{n+1} = l_n + 1$. The target is assumed to be detected when $l_n > 6$. The criterion of beginning of the signal is $l_n = 1$ and that of end of the signal region is $l_n > 6$, $l_{n+1} = 0$. The target azimuth is determined as the half-sum of the azimuths corresponding to the beginning and end of the signal region.

b) Sequential-observation type criterion. At the initial position the counter is set at zero. If at a certain position a one occurs, 4 is added to the counter; if a zero occurs one unit is subtracted. It is assumed that if $l_n \geqslant 16$, then $l_{n+1} = 0$ and the target is considered detected. If $l_n \leqslant -4$, then $l_{n+1} = 0$ and a target-absent signal is generated. The target region is measured from the first pulse obtained from the target to the first pulse indicating absence of target.

To determine the performance of both criteria we have to evaluate the detection probability and the false-alarm probability, and also find the distribution of the measurement errors of target azimuth (from which the error expectation and variance are found). The detection probability (and the false-alarm probability) for the series type criterion can be determined analytically by solving a system of recurrent equations with respect to the quantities p_n^k, denoting the probability that at the instant n the value k is recorded in the counter.

The error distribution cannot be evaluated by elementary methods. The use of the Monte Carlo method enables us to solve both problems simultaneously.

Electronic computers are usually employed for the solution of these problems. Locations of the machine memory are allocated for storing the content of the positions of an azimuth scan: ones, if signal (useful or noise) is present, and zeros in the opposite case. The method is implemented in the following manner: K locations, allocated for azimuth scan, are filled with zeros or ones according to a binomial distribution with parameter p_{N} in a noise region and parameter p_{S} in a signal region.* The realization which is obtained of the observed scan is analyzed by means of the corresponding criterion, i.e., if the presence of target is recorded, then its azimuth is determined and is compared with the true one. The value of the error is entered in a histogram. By repeating the simulation process many times we can determine the required characteristics to any desired accuracy. The correctness of the calculation may be partially checked by comparing experimental detection probabilities with theoretical ones obtained by solving a system of recurrent equations for series type criteria. Such a check will provide evidence of the quality of the binomial-distribution model.

Calculations of detection characteristics by these methods ordinarily require the execution of about 1000 tests for a single version (see [48b]).

Estimation of the Azimuth by the Maximum-Likelihood Method

The joint probability distribution of the sequence of voltage amplitudes at the receiver output depends, for a fixed distance, on the target azimuth α, i.e., can be written in the form $p(r_1, \ldots, r_n, \alpha)$. Therefore, the problem of measuring the target azimuth may be formulated as that of determining the unknown parameter of a distribution. A method for solving this problem, called the maximum-likelihood method, has been developed in statistics.

---

*We assume that a signal region is bounded on both sides by noise regions.

In this method the adopted estimate for the unknown parameter is that value $\hat{\alpha}$ of the parameter which maximizes $p(r_1^{\text{ob}}, \ldots, r_n^{\text{ob}}, \hat{\alpha})$. Here $r_1^{\text{ob}}, \ldots, r_n^{\text{ob}}$ are the observed values of the corresponding random quantities (in our case the voltage amplitudes). We shall assume that the radar receiver comprises a threshold device, which quantizes the signal into two values, and that the signal itself is noiselike, i.e., its amplitudes are Rayleigh-distributed and uncorrelated. Then the probability of occurrence of the kth pulse in a "packet" of reflected signals is equal to

$$p_k = \exp\left\{-\frac{x_0^2}{2[1+q^2 g(\alpha-\theta_k)]}\right\}, \tag{5.30}$$

where x_0 is the quantization level, q is the signal-to-noise ratio, $g(\alpha)$ is the normalized $[g(0)=1]$ radiation pattern of the antenna, θ_k is the azimuth of the kth pulse, and α is the true target azimuth. We shall again denote by μ_k the result of quantization of the amplitude in the kth position, which can be either one or zero. The equation for finding the target azimuth by the maximum-likelihood method has the form

$$\sum_{k=1}^{n} \mu_k \eta(\hat{\alpha}-\theta_k) = 0, \tag{5.31}$$

where

$$\eta(x) = \frac{g(x)\,g'(x)}{[1+q^2 g^2(x)]\left[1-\exp\left\{-\frac{x_0^2}{2(1+q^2 g^2(x))}\right\}\right]}$$

are weighting coefficients depending on the estimate $\hat{\alpha}$. In order to find the required azimuth estimate, we must move the set of "weights" $\eta_k = \eta(\hat{\alpha}-\theta_k)$ along the sequence of zeros and ones obtained as a result of quantization of the signals in the considered sector of the scan, i.e., having affixed the weights $\eta_1, \ldots, \eta_n$ to the extreme left elements of the scan, we build up the sum of those weights against which there is a one, and then move the weights by one position to the right and again find the weighted sum, etc., until a minimum (close to zero) of the weighted sums $\sum_{k=1}^{n} \mu_{k+l}\eta(\hat{\alpha}-\theta_k)$ is attained. The azimuth of the position corresponding here to the weight $\eta(0)$, i.e., having argument zero, is taken as the target azimuth.

To calculate the efficiency of the maximum-likelihood method, it would be necessary to find the probability distribution of the estimate $\hat{a}$. The analytical solution of this problem is clearly faced with great difficulties, and therefore one naturally has recourse to the Monte Carlo method. Calculations by the Monte Carlo method proceed according to the following scheme: azimuth scanning is simulated by means of a sequence of zeros and ones,* then weights $\{\eta_k\}$ are "superimposed" on this sequence, and the root of Eq. (5.31) is found and is entered in a histogram. It is shown in [48b] that 100 such tests sufficed to construct the probability distribution function to sufficient accuracy.

We observe that by the same method we can determine the azimuth-measurement accuracy in the case of other types of signal fluctuation.

The Probability Distribution of Noise Power at the Output of a Radio-Relay System ([54, 55])

The performance reliability of a communication radio-relay system is determined by the total intensity of the noise voltages arising in various sections of the system between transmitter and receiver stations. Therefore, the correct design of such a system requires a knowledge of the statistical characteristics of the total noise and, in the first instance, of the probability distribution of noise power at the system output. Of special interest is the behavior of the distribution function for large values of the argument (for engineering purposes it is required to estimate the threshold noise power that is exceeded with probability equal to an assigned small number $p_{\text{thr}} = 0.01;\ 0.001$).

The mathematical problem is formulated in the following manner: a sequence of random quantities $\xi_1, \xi_2, \ldots, \xi_n$ with distributions $F_1(x), F_2(x), \ldots, F_n(x)$ is given. It is required to find the distribution function of the sum

$$\xi_1 + \xi_2 + \ldots + \xi_n.$$

Owing to the large number of terms in it, it would be natural to use limit theorems of probability theory. However, this, at first sight obvious, approach is faced in a number of cases with serious computational difficulties. This is explained by the fact that the limiting distribution is found sometimes not to be normal. We shall restrict ourselves to the simple case when the random quantities ξ_k are independent and identically distributed.

---

*The probability of occurrence of a one in the kth position is determined by formula (Section 30); the probability of a noise is $p_{\text{N}} = e^{-\frac{x_0^2}{2}}$.

According to experimental data the distribution functions are of the form

$$F(x)=\begin{cases} 1-\frac{\varphi(x)}{x^{\alpha}}, & x^{\alpha}\geqslant\varphi(x), \\ 0, & x^{\alpha}<\varphi(x), \end{cases} \tag{5.32}$$

where $\alpha\,(1<\alpha<4)$ is a parameter and $\varphi(x)\to c\,(>0)$ for $x\to\infty$. It is shown in probability theory that the limiting distribution of the sum of independent random quantities distributed according to one and the same law (5.32) proves to be normal only in the case when $\alpha\geqslant 2$. It is evident that the calculation of the threshold power is in this case elementary. Difficulties occur when the parameter α satisfies the inequality $\alpha<2$. The limiting distributions are then found to belong to a class of stable distributions of which is known only the form of the characteristic function

$$f_{\alpha}(t)=\exp\left[i\gamma t-c|t|^{\alpha}\left\{1+i\beta\frac{t}{|t|}\omega(t,\alpha)\right\}\right],$$

where $|\beta|\leqslant 1$, $c\geqslant 0$ and

$$\omega(t,\alpha)=\begin{cases} \tan\frac{\pi}{2}\alpha, & \alpha\neq 1, \\ \frac{2}{\pi}\log|t|, & \alpha=1. \end{cases}$$

We can easily write down the distribution function by representing it as the Fourier transform of the characteristic function. However, the calculation of the probabilities of large deviations requires the evaluation of integrals of very rapidly oscillating functions. In this case, even by using electronic computers, it is difficult to obtain a satisfactory answer.

The use of the Monte Carlo method enables us to obtain the required results by a reliable, standard method: a sequence of n random numbers $\xi_1, \xi_2, \ldots, \xi_n$ distributed according to (5.32) is generated and the sum

$$\xi=\xi_1+\ldots+\xi_n$$

is entered in a histogram. According to [54] and [55], to ensure an acceptable accuracy in the determination of the distribution function as regards its tails (we have in mind arguments x_0 for which $p\{\zeta > x_0\} = p_{\text{threshold}}$) not less than 5000 tests are required.

The probability distribution of the components is assumed most often to have the form of Pareto's law

$$F(x) = \begin{cases} 1 - \dfrac{1}{x^\alpha}, & x \geqslant 1, \\ 0, & x < 1. \end{cases}$$

If the noise voltages of adjacent sections are not identically distributed (and are possibly correlated), then by introducing appropriate modifications in the procedure for simulating the random quantities, we can find the probability distribution of their sum ζ. We can similarly consider also more complicated functions of $\xi_1, \ldots, \xi_n$.

Methods of Decoding Messages [8]

Before a message is transmitted, it has to be coded. At the receiving end of the line, to determine the content of the signal received, the latter has to be suitably processed (decoded). The coding may be carried out in the individual letters of an explicit message (that is, each "word" formed by n code symbols denotes one of the letters of the alphabet). In other cases "words" of n code symbols are complete messages (thus it is possible to transmit γ^n different messages, if γ is the number of values a symbol can assume). Noise present in the communication line and in the receiver distorts the code symbols, and therefore the reconstruction of the initial message is liable to be in error. This gives rise to problems in the design of noiseproof codes and in the investigation of reliable methods of decoding.

We shall restrict ourselves to binary codes, the symbols of which assume only two values: 0 and 1. These symbols are transmitted by two electric signals $S_0(t)$ and $S_1(t)$ differing from each other in one or several parameters (amplitude, duration, etc.). Owing to noise in the communication channel and in the receiver, the signals at the receiver output, $z(t)$, do not coincide either with $S_0(t)$ or with $S_1(t)$. Each received signal can be identified by some method or other with one of the possible signals: either S_0 or S_1. In order to detect and correct errors we add, to n-symbol code "words" carrying useful information, k more check symbols. In Wagner's code this is done in the following manner: the sum of all

n symbols in a word is calculated; if the sum is even (odd), a zero (a one) is put in the $(n+1)$th position. Therefore, in each received $(n+1)$- symbol word the sum of the symbols must be even. If it is found that in the message received this condition is not satisfied, a single error can be corrected, for example, in the following manner. On the basis of known statistical properties of noise we calculate, for each symbol in the word, $p_0 = p(S_0/z)$ and $p_1 = p(S_1/z)$, i.e., the a posteriori probabilities of the occurrence of S_0 and S_1, with the condition that the signal z has occurred at this position. We assume that the original signal is $S_i\,(i=0,\ 1)$ if the a posteriori probability p_i corresponding to it is the larger.* For each position we work out the difference $\Delta p = p_0 - p_1$. We decide on the position of the wrong symbol in the "word" received (when the latter has odd parity) by considering the set of differences, Δp, and finding the position for which $|\Delta p|$ is a minimum. It is thus possible to correct, with a certain probability, single errors.

Other methods of coding and identifying the received signals with transmitted messages are known. The quality of a coding method is estimated by the probability of error in the reception of a message. These probabilities can be compared by simulating on a computer the transmission of "words," by using pseudo-random numbers representative of transmitted symbols distorted by channel noise (to do this the probability distribution of noise must be assigned). By repeatedly simulating reception by various methods, we can calculate the frequency of correct identification of the received message for each of the methods compared. In the case when the symbols are independent, it suffices to carry out the simulation with one "word" only (for example, the word consisting of all zeros).

---

*We can proceed in this manner if all combinations of code symbols are permissible and there is no correction between words.

Chapter VI

Generating Uniformly Distributed Random Quantities by Means of Electronic Computers

1. COMPARISON OF VARIOUS METHODS OF GENERATING RANDOM QUANTITIES

The success of calculations by the Monte Carlo method on an electronic computer is determined by two basic factors:

a) the quality of the source of random numbers,

b) the choice of a rational computing algorithm.

The problem of choosing the method of generating random numbers is one of paramount importance, since its successful solution decides in many cases the success of the solution of the whole problem.

The simulation of a given random process requires the possibility of sufficiently economical construction of sequences of random numbers corresponding to some assigned distribution laws. We have observed that to obtain a value of a random quantity with an assigned distribution law, one or several values of uniformly distributed random numbers are usually employed. Therefore the problem of computer generation of uniformly distributed random numbers is of particular importance.

The latter problem can be solved by various methods.

The first method, one that is widely employed today, consists in the following. Random numbers are obtained in a computer by programming some recurrence relation. This means that each successive number α_{j+1} is generated from the previous one α_j (or from a group of previous numbers) by employing some algorithm consisting of arithmetic and logical operations. Such a sequence of numbers, without being random, may nevertheless satisfy various statistical criteria of randomness. Therefore, such numbers are called pseudo-random.

The main advantages of the programming method of obtaining random numbers are, among others, the possibility of checking computer operation during the process of solving a problem (the possibility of repeatedly reproducing a calculation) as well as the simplicity of the algorithm for generating a pseudo-random number.

The main shortcoming of pseudo-random numbers is the difficulty of a theoretical assessment of their statistical properties. This is particularly evident when various multidimensional distributions are generated by proceeding from a sequence of uniformly distributed pseudo-random numbers. In addition, all sequences of pseudo-random numbers generated by programming methods are periodic, and therefore, even from a practical point of view, very long sequences will not be random. Several methods of ''improving'' sequences of pseudo-random numbers are available today and will be discussed in detail below. However, the use of these methods can somewhat reduce the speed of operation of the computer.

Methods of obtaining uniformly distributed pseudo-random numbers are described in Section 2.

The second method of obtaining random numbers consists in the use of a special computer accessory, namely a random-number generator, which transforms the results of a physical random process into a sequence of binary digits in the computer, i.e., generates a random quantity. Usually, the register where this random quantity is located has an address in the general system of addresses of the computer. In that case, access to the random-number generator is obtained simply by referring to the computer memory.

The use of a random-number generator increases the speed of computation, since at each step of operation of the computer a new random number is written in a fixed standard location of the memory.

A disadvantage of such a method is a certain instability of the random-number generator owing to which it needs periodic checking. Another disadvantage is the impossibility of exactly reproducing the results of computation of a problem (computations cannot be checked by performing them twice).

It is usually reckoned that if the Monte Carlo method is employed systematically on a given computer, then the presence of a checked-up random-number generator is to be preferred to the use of pseudo-random numbers. If, however, the Monte Carlo method is employed on a computer only occasionally, then the use of programming methods is more advantageous, since the operation of random-number generators and their maintenance require considerable expenditure of labor. The design and operation of random-number generators are described in Section 4.

A third, comparatively rarely used, method consists in inserting tables of uniformly distributed random numbers into the working memory of the computer.

The use of a part of the working memory of a computer for storing tables of random numbers is usually impossible since the capacity of the computer memory is already utilized for storing information of various types directly related to the process to be simulated. As to the insertion of uniformly distributed random numbers into the working memory of the computer by means of drums or tape, the longer access times of such low-speed memory devices substantially reduce the speed of operation of the computer. In addition to this, in order to solve large problems, hundreds of thousands or even millions of random numbers are often required, and this exceeds by many times the volume of tables of uniformly distributed random numbers available today.

All that has been said above leads to the conclusion that the method of inserting uniformly distributed random numbers in the computer memory can be of only auxiliary value. This method is usually employed in the solution of small problems.

2. OBTAINING UNIFORM PSEUDO-RANDOM NUMBERS ON COMPUTERS

It has already been pointed out above that by "sequences of pseudo-random numbers" we mean sequences of values α_i of functions of an argument i (an integer) that resemble, in their properties, sequences of random numbers and satisfy a determinate system of accepted statistical criteria. Pseudo-random uniformly distributed numbers are generated on computers by means of special programs. The designing of a program to compute uniformly distributed pseudo-random numbers must take into account the following basic requirements.

a) The program must generate numbers having very little statistical correlation. A sample set of pseudo-random numbers generated by the program must satisfy prescribed criteria of "randomness."

b) The distribution of pseudo-random numbers generated by the program must approximate as closely as possible to a uniform distribution.

c) The program must be stable. This means that the distribution of the random numbers generated must not vary during the operation of the program.

Let us consider some of the methods available for obtaining uniformly distributed pseudo-random numbers.

1. Analytical Methods of Obtaining Pseudo-Random Numbers

a) A whole series of methods of obtaining uniformly distributed pseudo-random numbers is based on selecting the central digits of a product. This method, which was firstly proposed in 1951 by J. von Neumann [223], consists in the following. An arbitrary number α_0 consisting of $2k$ binary digits is selected to start a recurrent process. The quantity α_0 is squared. The quantity α_0^2 consists of $4k$ digits from which a number α_1 is chosen, consisting of the $2k$ central binary digits [from the $(k+1)$th to the $(3k)$th]. Then the process is repeated starting from α_1, etc.

Such a type of recurrent process does not give a satisfactory sequence of random digits, and the distribution of the pseudo-random numbers obtained by this method differs from a uniform distribution.

Considerably better results are obtained by a modified version of von Neumann's method, in which a pair of numbers, for example α_0 and α_1, is arbitrarily chosen. Their product $\alpha_0\alpha_1$ is formed and its central digits are used for the number α_2. The process is repeated for α_1 and α_2 to give α_3, etc. This type of recurrent process gives pseudo-random numbers with a distribution closer to the uniform distribution than does von Neumann's "mid-square" method in its original version.

A series of similar methods for obtaining uniform pseudo-random numbers has been suggested by Lehmer [189]. They consist in the following. Firstly an initial number α_0 is chosen consisting of $2k$ figures (of $2k$ binary digits on an electronic computer). The product α_0^2 is formed and the last $2k$ figures of this product are chosen, yielding a new number α_0'. The number α_0' is multiplied by a constant factor C and the first $2k$ figures of this product are selected to give a number α_0''. The first $2k$ figures of the product α_0^2 are then selected and the number α_0''' is formed. The number α_0''' is multiplied by C and the last $2k$ figures of the product $C\alpha_0'''$ are selected to give the number α_0^{IV}. Digit-by-digit (or ordinary) addition of the numbers α_0'' and α_0^{IV} gives the next pseudo-random number α_1.

The methods described above (also termed truncation methods of pseudo-random numbers) give in the end a periodic sequence with period not exceeding 2^{2k}.

b) A series of methods of obtaining pseudo-random numbers is based on the use of residues. On the ENIAC computer* Lehmer [189] has used the recurrence relation

$$\alpha_{n+1} = k\alpha_n \pmod{M},$$

where $k = 23$, $M = 10^8 + 1$.

---

*The ENIAC computer uses the decimal number system.

Such a method enables us to obtain a sequence of eight-figure decimal numbers with period 5882352.*

In the paper [251], pseudo-random numbers uniformly distributed over the interval (0, 1) have been generated by using the following recurrence relation

$$\alpha_{j+1} = 2^{-42}\beta_j,$$
$$\beta_{j+1} = 5^{17}\beta_j \pmod{2^{42}}, \quad \beta_0 = 1.$$

Such a set has a period equal to $2^{40} \sim 10^{12}$.

The set of pseudo-random numbers generated by means of the recurrence relation [247]

$$\alpha_{j+1} = 2^{-36}\beta_j,$$
$$\beta_{j+1} = 5^{13}\beta_j \pmod{2^{36}}, \quad \beta_0 = 1,$$

has period $2^{34} \sim 2 \times 10^{10}$, while the set obtained from the recurrence relation

$$\alpha_{j+1} = 10^{-10}\beta_j,$$
$$\beta_{j+1} = 7\beta_j \pmod{10^{10}}, \quad \beta_0 = 1,$$

has period 5×10^7.

The first method is used on the SVAC computer and the second on the OARAC computer.

On the UNIVAC electronic computer a recurrence relation of the following form [212] has been used

$$\alpha_0 = 1, \quad \alpha_{n+1} = 10^{-11}\beta_n,$$
$$\beta_{n+1} = 7^{4k+1}\beta_n \pmod{10^{11}}.$$

The resulting sequence of pseudo-random numbers has period 5×10^8.

A number of methods for generating uniformly distributed pseudo-random numbers have been developed, using a Fibonacci type series. For example, good results have been obtained by using the sequence

---

*The length of a period has been calculated by an empirical method.

$$F_0 = 0, \quad F_1 = 1,$$
$$\alpha_{n+2} = \frac{1}{M} F_{n+2},$$

where

$$F_{n+2} = [F_{n+1} + F_n] \pmod{M}.$$

For the SEAC computer, $M = 2^{44}$ has been used. The period thus obtained was approximately equal to 2.5×10^{13}.

Van Weingarden [251] has suggested a modified version of the last-mentioned method. A group of numbers $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_p$ is arbitrarily selected, and a linear combination of them with integral coefficients is formed. Its fractional part is the number α_{p+1}. The process is repeated on the set $\alpha_1, \alpha_2, \ldots, \alpha_p, \alpha_{p+1}$, etc. If the numbers $\alpha_0, \ldots, \alpha_p$ are uniformly distributed over the interval (0, 1), then, from the elementary theorem stating that the sum of two uniformly distributed quantities, taken modulo 1, is uniformly distributed, it follows that α_{p+1} is also uniformly distributed. It can easily be shown for $p = 1$ that any two quantities of this sequence are independent. This cannot be said of triplets, since they are found in certain cases to be linearly dependent.

With such a scheme we obtain longer periods, since, before repetition begins, the whole set of $p+1$ numbers must again assume its initial values.

We observe that in obtaining pseudo-random uniform numbers generated according to the formula

$$\alpha_{n+j} = \sum_{k=1}^{j-1} C_k \alpha_{n+k} \pmod{M},$$

the choice of the initial values $\alpha_0, \alpha_1, \ldots, \alpha_{j-2}$ plays an important role. Thus, with a suitable choice of the first two initial quantities α_0, α_1, it has been possible to obtain on a SEAC computer a sequence of pseudo-random numbers satisfying a system of various statistical criteria [105]. Pseudo-random numbers are generated according to the formula

$$u_{j+1} = [u_j + u_{j-1}] \pmod{4},$$

where

$$u_0 = \pi, \quad u_1 = 5^{17} \cdot 2^{-42}.$$

c) Several procedures for obtaining uniformly distributed pseudo-random numbers are based on Lehmer's method [251].

The latter relies on the known fact that for an irrational number θ the set of numbers of the type $\{n\theta\}$ (where $\{x\}$ denotes the fractional part of the quantity x) comprises numbers arbitrarily close to any number in the interval (0, 1), since they are everywhere dense in (0, 1). Lehmer's method reduces to constructing a sequence $\alpha_n = \{n\theta\}$ of nonrandom correlated numbers the distribution of which (in the number-theoretical meaning of this word) is rigorously uniform. This method is used with advantage in those cases where the main interest is the uniformity of the distribution while the presence of correlation is of little importance.

Various schemes based on number-theoretical methods have been suggested for generating pseudo-random sequences by N. M. Korobov and I. I. Shapiro-Pyatetskii [39].

II. Obtaining Uniform Pseudo-Random Numbers by the Mixing Method.

For obtaining uniform pseudo-random numbers on the electronic computers "Strela," BESM, "Ural," etc., several methods exist that utilize specific features of these machines. These methods are based on simulating the random, chaotic mixing of the digits of the mantissa of pseudo-random numbers. Among these methods the most fully developed are those for "Strela" type electronic computers. We shall cite, as typical of such programs, those for generating pseudo-random numbers for the "Strela" computer developed by I. M. Sobol' and D. I. Golenko. Both these recurrent methods are realized by means of a three-instruction program.

In the program developed by I. M. Sobol' [57], the number α_{k+1} is obtained from the number α_k by three operations:

1) the number α_k is multiplied by 10^{17};

2) the number representing the product $10^{17}\,\alpha_k$ (the "image" of $10^{17}\,\alpha_k$) is shifted seven places to the left (so that zeros will be found at places from 36 to 42);

3) the absolute value of the number obtained is taken (and at the same time the number is normalized). The result is α_{k+1}.

We observe that the properties of the sequence $\{\alpha_k\}$ depend not only on the value assigned for α_0 but also on the method of rounding off used in the computer.

Moreover, the number 10^{17} and the number for determining the value of the shift are placed in the unit for issuing the constants of the computer. Thus, only one location of the working memory is required to compute α_k.

Finally, the risk of degeneracy of the sequence, that is the probability of some α_k reducing to zero, is extremely small in practice. In fact, for α_k to reduce to zero it is necessary that

$$0.5 \leqslant \alpha_{k-1} \cdot 10^{17} < 1;$$

otherwise, nonzero digits will occur in the exponent of the product $\alpha_{k-1} \times 10^{17}$, and after the subsequent shift these digits will be found in the mantissa of α_k. Thus, in order that the sequence may degenerate, α_{k-1} must belong to a very narrow interval of values.

Several tens of values of α_0 have been used in the solution of various problems, but no cases of degeneracy of the sequence have occurred.

The three-instruction program developed by D. I. Golenko [24] has also been repeatedly used in the solution of a series of problems by the Monte Carlo method.

Pseudo-random numbers uniformly distributed over the interval (0, 1) are generated by three single-step computer operations. An initial assigned value $\alpha_0 \neq 0$ (which can be also random) is placed to begin with in the location *a*, in which the subsequent numbers of the sequence are to be generated.

Pseudo-random numbers are calculated by recurrence; the number α_{k+1} is generated from α_k according to the following program.

1. The image of the number α_k in the location *a* is shifted seven places to the left (the places 36 to 42 are "cleared") and is placed in the location *b*.

2. The special addition operation* (the operation 02) is carried out on the contents of the locations *a* and *b*, the first number added being the content of the location *b*. As a result, zeros arise in the places from 36 to 42, while the content of the places of the mantissa of the locations *a* and *b* is chaotically mixed. The result is written in the location *a*.

3. The absolute value of the content of the location *a* is evaluated (with subsequent normalization). The result is written again in the location *a* (the operation 04). This result is the number α_{k+1}.

In the instruction-code language of the "Strela" computer this program will be written thus

	Operation code	1st address	2nd address	3rd address
1	14	*a*	4007	*b*
2	02	*b*	*a*	*a*
3	04	*a*	—	*a*

---

*The special-addition instruction accomplishes the logical digit-by-digit addition of the contents of the first and second addresses, the exponent of the first number added being retained. The first is written in the third address.

Thus two locations of the accumulator, a main one (a) and an auxiliary one (b), are needed for obtaining pseudo-random numbers. In the intervals between the generation of pseudo-random numbers the location b may be used as a working location.

Just as for the pseudo-random number program described above, the probability of degeneracy of the number sequence generated is very small in practice. Degeneracy occurs only in the case when the content of the location b after the shift is the digit-by-digit complement of the content of the location a. Only in this case does the special addition operation result in a zero in the location a. Cases of degeneracy of the sequences have never been observed, in spite of the fact that pseudo-random numbers have been used for several years in a large number of problems with various α_0.

The first of the programs described above has an aperiodicity stretch (see Section 3) of the order of 80,000, the second program one of the order of 300,000. The length of the cyclicity period generated amounts to 50,000 numbers for the first program and to about 100,000 numbers for the second one.

The program for obtaining on the BESM computer pseudo-random numbers uniformly distributed over the interval (0, 1) comprises nearly the same machine operations as the corresponding program for the "Strela" computer. The following table gives the four-instruction program for the BESM computer [50]

Ordinal number of instruction	Operation code	Address			Operation content
		I	II	III	
k	, ←	a	0047	$a+1$	7-digit right shift
$k+1$	, ←	a	0007	$a+2$	7-digit left shift
$k+2$	SA	$a+1$	$a+1$	$a+2$	Special addition
$k+3$	|TM|	$a+2$	0000	a	Transfer of modulus of number with normalization

This program differs from the corresponding program for the "Strela" computer owing to the different computer-operation codes for the "Strela" and the BESM. Thus the second shift in the program for the BESM computer is necessary since without it there occurs degeneracy of the pseudo-random number program with consequent "zero output."

The length of the aperiodicity stretch for this program does not exceed 50,000. As to the length of the period of the sequence of pseudo-random numbers, this is close to 5000 for several initial values of α_0. In the process of solving certain problems by the Monte Carlo method, several modifications have been introduced in the program for the BESM computer: the number of positions shifted was increased to 8, etc. This has increased the length of the aperiodicity stretch to 100,000-150,000, but the quality of the pseudo-random numbers thus generated has appreciably deteriorated [50].

As to the single-address electronic computer "Ural-1," the program for obtaining on it pseudo-random numbers uniformly distributed over the interval (0, 1) is more cumbersome than the programs described above.

Provided they are taken in sufficiently long sequences, the pseudo-random numbers obtained by means of the above-described programs usually satisfy the system of statistical criteria adopted for testing the randomness of the distribution. These criteria will be described in detail in Section 3. However, the use of statistical goodness-of-fit criteria does not solve the problem of a thorough assessment of the distribution of pseudo-random numbers. The fact that a statistical criterion is obeyed merely means that we have no grounds for rejecting the hypothesis of uniformity of the distribution of the pseudo-random numbers. This, however, does not by any means guarantee that this hypothesis is the correct one.

In addition to this, in a series of statistical goodness-of-fit criteria (for example the χ^2 test) the closeness of the empirical distribution to a uniform distribution may have to be variously assessed depending on the sample size n. Therefore, for one and the same empirical distribution we may accept the hypothesis of a uniform distribution for one value of n and reject it for another value of n.

What has been said above does not detract from the importance and usefulness of the given statistical goodness-of-fit criteria for assessing the empirical distribution of a sequence of pseudo-random numbers. Goodness-of-fit criteria are a very effective method of analysis of the quality of pseudo-random numbers. This is especially true in those cases in which not one, but a system of statistical criteria supplementing and reinforcing each other is used. We shall merely remark that in a number of concrete cases, just as with other methods of statistical analysis, the use of goodness-of-fit criteria can be rather inefficient or altogether too difficult to carry out.

The degree of closeness of an empirical distribution of pseudo-random numbers to the theoretical uniform distribution may also be investigated without the use of statistical goodness-of-fit criteria, if we analyze the distribution of the separate digits of which the pseudo-random numbers consist.

We shall assume below that the pseudo-random numbers in question have been obtained by means of the programs described above from an initial random number α_0. Otherwise we would not be entitled to apply to pseudo-random numbers such probability-theoretical concepts as randomness, probability, distribution, etc.

Let us consider the distribution of the values of separate digits for several pseudo-random numbers generated from each other (i.e., obtained sequentially). If a pseudo-random number obtained is represented in the form of a nonnormalized n-digit binary number with zero exponent, then, in the case of uniform distribution, the probability of occurrence of a zero (or one) at an arbitrary place of the mantissa is equal to 0.5. In reality, for each position, this probability is somewhat different from 0.5.

Suppose a sequence of pseudo-random numbers is generated and let the symbol $\xi_i^{(k)}$ denote the ith digit in the kth calculated pseudo-random number, α_k. Then the probability of occurrence of a zero in $\xi_i^{(k)}$ may be written in the form $0.5(1+\delta)$.

Let us consider the connection between the distribution of the values of the separate digits of nonnormalized pseudo-random numbers and the probability distribution of pseudo-random numbers themselves. Let us introduce the notation

$$p = \max_{1 \leqslant i \leqslant n} \mathrm{P}\{\xi_i^{(k)} = 0\} = 0.5(1+\delta_{\max})$$

and let us estimate the relative deviation Δ of the distribution of a sequence of pseudo-random numbers (in which all digits of the mantissa have a "probability distortion," equal to $\delta_{\max}$) from the uniform distribution. We shall disregard the correlation between digits and all digits of the mantissa will be assumed independent. Denoting by m the number of zeros in a pseudo-random number, we can easily convince ourselves that

$$\Delta = (1+\delta_{\max})^m (1-\delta_{\max})^{n-m} - 1 = (2m-n)\delta_{\max} + o(\delta_{\max}).$$

Therefore, for $m \sim \frac{n}{2}$

$$\Delta \sim 0,$$

and for $m=0$ or $m=n$

$$|\Delta| \sim n\delta_{\max}.$$

If we fix the number $\Delta_0 (1 > \Delta_0 > 0)$ and set ourselves the task of finding $\delta_{\max}$ for which $|\Delta| < \Delta_0$, then

$$\delta_{\max} < \frac{\Delta_0}{n}.$$

We have thus shown that for any preassigned Δ_0 there exists a value $\delta_{\max}$ such that if $\delta < \delta_{\max}$ for all digits of the mantissa of pseudo-numbers, then $|\Delta| < \Delta_0$ for the random numbers themselves. It remains to establish the value $\delta_{\max}$ for digits of pseudo-random numbers obtained by means of the programs described above, and thus to guarantee a relative departure from the uniform distribution not exceeding the value $n\delta_{\max}$.

It will be recalled that the last instruction in all three programs described above for obtaining uniformly distributed pseudo-random numbers (for the "Strela" and BESM computers) is the instruction for taking the modulus with subsequent normalization. Since the investigation of the distribution of the digits of a pseudo-random number α_{k+1} after its normalization is made very difficult by the deterministic nature of individual digits (the value of the first digit, for example, is always one, and that of the last is always zero), it is expedient to assess the value of $\delta_{\max}$ for digits of the mantissa of the pseudo-random number before its normalization.

As has already been indicated above, the initial number α_0 from which the sequence of pseudo-random numbers is generated (this initial number is not used as a pseudo-random number) is taken to be a random number with independent, random and uniformly distributed digits.* We can easily convince ourselves that $\delta_{\max} = 0$ for the first pseudo-random number α_1. The calculation of the value of $\delta_{\max}$ for α_k $(k \geqslant 2)$ is considerably more complicated. Although all calculations have been carried out only for the second of the above-described three-instruction programs for generating pseudo-random numbers on the "Strela" computer, the procedure suggested for calculating $\delta_{\max}$ may be employed also for other methods of obtaining pseudo-random numbers.

Calculations carried out have shown that, for digits from the first to the twentieth, the formula for computing the probabilities for $1 \leqslant k \leqslant 15$** has the following approximate form

$$|\mathsf{P}\{\xi_i^{(k)} = 0\} - 0.5| < 0.0001 \qquad (1 \leqslant i \leqslant 20).$$

---

*The number of digits is equal to the number of digits of the corresponding electronic computer.

**No calculations have been carried out for $k > 15$.

In other words, for the first twenty digits $\delta_{\max} < 0.0001$, and for $n = 20$ we have $\Delta < 0.002$.

It is important to note the following. The calculations carried out guarantee the uniformity of the distribution only of the first twenty (out of 35) digits of the mantissa of a pseudo-random number. This means that the probability distribution of the whole pseudo-random number may have periodic departures from uniformity with period not exceeding 2^{-20}.

Thus an analysis of the distribution of the values of the digits of several sequentially obtained pseudo-random numbers guarantees uniformity to 0.5% accuracy over the interval (0, 1) for at least 15 sequentially obtained pseudo-random numbers.

If the solution of a problem by the Monte Carlo method utilizes a sequence of pseudo-random numbers which have been given a preliminary check by means of a system of statistical tests, it is expedient to use just one program of uniform pseudo-random numbers without any improvement whatsoever. Otherwise a random "perturbation" of the digits of the pseudo-random number must be carried out periodically every 10-15 sequentially generated pseudo-random numbers.

Good results are obtained, for example, by using at the same time two programs in the following manner: at first an initial number α_0 is chosen and is written in the location a, and the generation of pseudo-random numbers begins according to the first program. After 10-15 pseudo-random numbers have been generated, we transfer momentarily to the second program, which generates, from the last pseudo-random number (the content of the location a), a new pseudo-random number which is used to start again the first program, etc. Such a use of two programs enables the length of the period to be greatly increased.*

We proceed now to a description of a system of statistical criteria to test the quality of pseudo-random numbers.

3. CRITERIA TO TEST THE QUALITY OF UNIFORM PSEUDO-RANDOM NUMBERS

a) Criterion for Testing the Periodicity of a Sequence of Pseudo-Random Numbers

A sequence of pseudo-random numbers obtained by means of a program has the following properties. The first L sequentially obtained pseudo-random numbers, beginning with α_1, will be all different, but the $(L+1)$th pseudo-random number α_{L+1} coincides with one of the numbers previously obtained, say, α_l $(1 \leqslant l \leqslant L)$ (Fig. 29). Further on, the subsequence of pseudo-random numbers beginning with α_l and ending with α_L will be periodically repeated. The number L is referred to as the length of the aperiodicity stretch, and the number $(L-l+1)$ as the length of the period. Of course the length of the aperiodicity stretch and the length of the period will be different for various values of the initial quantity α_1.

---

*If "perturbation" is carried out every m numbers, the average length of the period increases, as has been shown by I. M. Sobol', by $\sqrt{m}$ times.

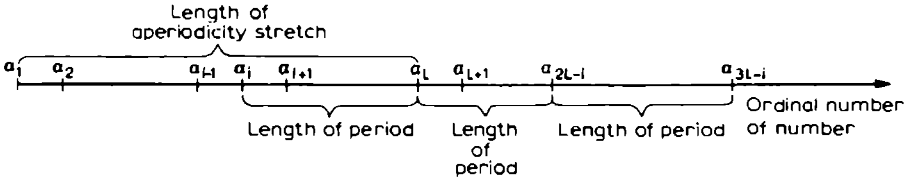


Fig. 29. Representation of sequence of pseudo-random numbers.

It must be pointed out that, in solving problems by the Monte Carlo method on computers, it is desirable that the number of pseudo-random numbers used be not larger than L. Otherwise the probability process would be simulated by means of repeating pseudo-random numbers, which leads as a rule to incorrect results.

Let us consider the problem of determining the length of the aperiodicity stretch.

In an electronic computer, only a finite number N of different pseudo-random numbers can be represented. For pseudo-random numbers distributed over the segment (0, 1) we have $N=2^l$, where l is the number of digits in the mantissa of the computer.

Suppose we had, instead of N pseudo-random numbers, N ordinary random numbers with equal probability of occurrence for any of them. In this case the length of the aperiodicity stretch L would be a random quantity the distribution of which would be assigned by the probabilities

$$\mathsf{P}\{L=x\}=\frac{N!\,x}{(N-x)!\,N^{x+1}} \qquad (x=1,\ 2,\ \ldots,\ N).$$

Let us consider the asymptotic distribution of the random quantity $\eta=\frac{L^2}{N}$ for $N\to\infty$. It may be shown that

$$\lim_{N\to\infty} \mathsf{P}\left\{\frac{L^2}{N} < z\right\} = 1 - e^{-\frac{z}{2}}.$$

The latter means that the random quantity $\eta = \frac{L^2}{N}$ is asymptotically distributed as χ^2 with two degrees of freedom (see Section 3c).

Let $\alpha_0^{(1)}, \ldots, \alpha_0^{(m)}$ be various initial values with which we begin to generate a sequence of pseudo-random numbers, and let $L_1, \ldots, L_m$ be the corresponding lengths of the aperiodicity stretch. By employing similar reasoning, it may be shown that the random quantity

$$\eta = \frac{\sum_{i=1}^{m} L_i^2}{N}$$

is distributed asymptotically as χ^2 with $2m$ degrees of freedom. The last fact enables us to use the χ^2 criterion for assessing empirical values of the length of the aperiodicity stretch.

Let us cite, for example, data obtained by generating pseudo-random numbers on the "Strela" computer by means of the program described on page 203. Calculations gave the following values of the aperiodicity stretch

$$L_1 = 2.3 \cdot 10^5, \quad L_2 = 3.5 \cdot 10^5, \quad L_3 = 2.8 \cdot 10^5.$$

In the "Strela" $N = 2^{35}$. Hence

$$\frac{L_1^2 + L_2^2 + L_3^2}{N} = 8.7.$$

In our example $m = 3$, which corresponds to a χ^2 distribution with six degrees of freedom. The value 8.7 does not exceed the 5% confidence limits for χ_6^2 (see Section 3c). Thus the experimental data of the example quoted indicate that, as far as the periodicity parameters are concerned, pseudo-random numbers simulate random numbers well.

b) Criteria for Testing Randomness

A system of tests designed to check the "randomness" of a distribution has been developed by Kendall and Babington-Smith. The system consists of four tests: the frequency test, the serial test, the gap test and the poker test.

A set of pseudo-random digits satisfying all these tests is called locally random ([157, 179]).

All the above-mentioned tests are characterized by one and the same general property: the pseudo-random numbers (or digits in them) subjected to a test are classified according to certain indices (different for each test), and the empirical distribution obtained is compared with the theoretical one. The χ^2 criterion, Kolmogorov's criterion and the ω^2 criterion are usually employed for the comparison.

Here is a description of Kendall's system of tests.

The frequency test involves counting the number of pseudo-random numbers that are found in subintervals of the range of definition of the pseudo-random numbers, i.e., the segment (0, 1). Usually the segment (0, 1) is subdivided into 10 to 20 equal intervals.

The frequency test is successfully employed also in the case of a multidimensional distribution. In this case we subdivide the multidimensional cube $0 \leqslant x_i \leqslant 1$, $1 \leqslant i \leqslant n$ into elementary small cubes with equal surfaces, and then we count the number of pseudo-random numbers that are found in each of the small cubes from a sample of a sufficiently large number of pseudo-random numbers. In practice, the size of the set of pseudo-random numbers investigated is rarely smaller than 10,000 and is usually 50,000-100,000. Of course, in investigating uniform pseudo-random numbers, the theoretical distribution to be compared with the empirical distribution is uniform, and the expected number of pseudo-random numbers in any interval is equal to $N \operatorname{mes} (V_l)$, where N is the volume of the sampling set, and $\operatorname{mes} (V_l)$ is the length of the subinterval (or the volume of the elementary small cube).

The serial test is the simplest one and involves counting the zeros or ones in the digits of the set of pseudo-random numbers for which the test is carried out. In the case of pseudo-random numbers uniformly distributed over the interval (0, 1) the expectation of the content of each place of the mantissa of a pseudo-random number is equal to $^1/_2$, since the probability of the occurrence of a zero or one is equal to 0.5.*

The "poker" test involves counting, in a sample set of large size, the number of pseudo-random numbers containing various combinations of binary digits. For example, let 14,000 pseudo-random numbers each containing 10 binary digits be subjected to testing. Table 11 shows an empirical distribution of the frequency of various combinations of binary digits in comparison with the theoretical uniform distribution.**

---

*This applies to nonnormalized pseudo-random numbers with zero exponent.

**Mathematical Tables and Other Aids to Computations, Vol. 101, No. 53 (1956).

Table 11

		Data of tests	Theoretically calculated data	
10 zeros		12	13.67	
9 zeros	1 one	149	136.72	
8 zeros	2 ones	607	615.23	
7 zeros	3 ones	1662	1640.63	$\chi^2 = 7.372$
6 zeros	4 ones	2922	2871.09	
5 zeros	5 ones	3468	3445.31	P = 0.69
4 zeros	6 ones	2763	2871.09	
3 zeros	7 ones	1633	1640.63	
2 zeros	8 ones	627	615.23	
1 zero	9 ones	143	136.72	
	10 ones	14	13.67	

The gap test is usually replaced by a frequency test of the length of runs [33].

This latter method applies to the case in which all pseudorandom numbers in the set investigated are divided into two classes: class a and class b.

A run is a section of the sequence consisting of numbers of the same class following each other. The number of numbers in a run is called its length.

In the theory of runs the following concepts are usually employed for testing "randomness":

r_{1i}—the number of runs of class a of length i,
r_{2i}—the number of runs of class b of length i,

$R_{1k} = \sum_{i=k}^{n_1} r_{1i}$ —the total number of runs of class a of length equal to or larger than k (n_1 is the maximum length of a run of class a),

$R_{2k} = \sum_{i=k}^{n_2} r_{2i}$ —the total number of runs of class b the length of which is not smaller than k,

$R_1 = \sum_{k=1}^{n_1} r_{1k}$—the total number of runs of class a,

$R_2 = \sum_{k=1}^{n_2} r_{2k}$—the total number of runs of class b,

$R = R_1 + R_2$–the total number of runs.

In checking "randomness" by the method of runs we must compare the theoretical and the empirical distributions of the quantities R_{1k}, R_{2k}, R_1, R_2 and R in a large sample of pseudo-random numbers. In this comparison the confidence probability p must be assigned.

To investigate pseudo-random numbers uniformly distributed over the interval (0, 1), the subdivision of the numbers into two classes is usually done in the following manner.

To the first class a are attributed numbers smaller than 0.5, to the second class b numbers larger than 0.5. The sampling values of the number of runs and of the length of runs of numbers of the first and second classes are compared with their theoretical limits determined according to the following formulas: for sufficiently large n (practically for $n \geqslant 20$) and $p = 0.95$ the lower limit of the total number of runs is equal to

$$z = \frac{1}{2}\left(n + 1 - 1.65\sqrt{n-1}\right)$$

(e.g., for $n = 10{,}000$ z equals 4918); the lower limit of the number of runs of numbers of the first (or second) class is equal to

$$z(R_1) = z(R_2) = \frac{1}{4}\left(n - 1.65\sqrt{n+1}\right)$$

(e.g., for $n = 10{,}000$ $z(R_1) = z(R_2) = 2459$).

As to the distributions of the quantities R_{1k} and R_{2k} and the determination of the limits of the length of runs, k, it can be shown by methods of combinatorial analysis that the upper limit of the length of runs is

$$z(k) = \frac{\lg\left(-\frac{n}{\ln p}\right)}{\lg 2} - 1$$

(for $n = 10{,}000$ and $p = 0.95$ we have $z(k) = 15$).

This means that a run of length 15 or larger (for each of the classes) may be expected to occur in a random sequence of 10,000 random numbers in only 5% of the cases.

Below we describe criteria used in comparing empirical with theoretical distributions.

c) Criteria for Testing Uniformity

A sequence of pseudo-random numbers is usually tested for "uniformity" on the basis of three criteria: the χ^2, Kolmogorov's and the ω^2 criterion.

The basic, most often used criterion is the χ^2 test.

The χ^2 test is based on the statistic

$$\chi^2=\sum_{i=1}^{l}\frac{(\nu_i-np_i)^2}{np_i},$$

where ν_i is the sample number of objects in the ith interval, and np_i is the expectation of the quantity ν_i in the hypothetical distribution (in our case the hypothetical distribution is the uniform distribution over the segment (0, 1).

If the segment (0, 1) is subdivided into l equal intervals, and if, having taken a sample of sufficiently large size n, we evaluate the sum

$$\frac{1}{ln}\sum_{i=1}^{l}(l\nu_i-n)^2 \qquad \left(p_1=\ldots=p_l=\frac{1}{l}\right),$$

then in the case of uniformity of the theoretical distribution this sum will approximately follow the χ^2 distribution with $(l-1)$ degrees of freedom.

The hypothesis of uniform distribution of the pseudo-random numbers must be rejected if χ^2 exceeds the upper limit $\chi^2_{l-1}(p)$ of the confidence interval, where p is the assigned confidence probability and $l-1$ is the number of degrees of freedom. By confidence probability p we mean a probability close to 1 such that, if the hypothesis of uniform distribution is correct, the probability is p that the value obtained for χ^2 will not exceed $\chi^2(p)$. If however the confidence limit is exceeded, this means that the measure of discrepancy, χ^2, indicates a significant departure, and the hypothesis of uniform distribution must be rejected.

Extremely small values of χ^2 must be considered an indication of failure of randomness, since in the case of correctness of the hypothesis the probability for a random quantity to assume too small values is extremely small. Therefore, the critical region of "nonacceptance of the hypothesis" is conveniently taken to have two parts. A lower limit for χ^2 may be established as follows:

$$\mathrm{P}\left\{\chi^2>\chi^2_{l-1}(p)\right\}=\mathrm{P}\left\{\chi^2<\chi^2_{l-1}(1-p)\right\},$$

where $\chi^2_{l-1}(1-p)$ denotes the lower confidence limit.

The fact that the value of χ^2 is found once in the interval $[\chi^2_{l-1}(1-p), \chi^2_{l-1}(p)]$ does not make it certain that the hypothesis is correct, just as the fact that χ^2 exceeds once the limits of the interval (especially for $p = 0.8 - 0.9$) does not make it certain that the hypothesis is false. In order to be certain of our judgment as to the hypothesis, we have to carry out the calculation of χ^2 several times and examine how closely the empirical distribution of the quantity agrees with the theoretical one. Thus, in the case of $p = 0.95$, the probability of lying once beyond the limits of the interval is small but far from negligible. In the case of correctness of the hypothesis, the value may be beyond the limits of the interval, or the average, in one case out of ten. However, if sampling is carried out twice and the values lie beyond the limits of the interval in both cases, then this event has probability 0.01 and must be considered practically impossible. Results of tests of pseudo-random numbers using the χ^2 criterion are shown in Table 13 on page 219.

If in multiple testing using Pearson's criterion the quantity χ^2 does not exceed the value $\chi^2_{l-1}(p)$, but each time differs from it little, then this is evidence of the need for carrying out additional investigations owing to the inadequate agreement of the hypothetical and the empirical distributions.

A very effective measure of goodness-of-fit is also Kolmogorov's criterion which is based on the statistic

$$D_n = \max |F_n(x) - F(x)|,$$

where $F_n(x)$ is the empirical distribution function, which is equal to $\tilde{p} = \frac{m_x}{n}$ (n is the size of the sample, m_x is the number of objects in the sample not exceeding the value x), while $F(x) = \mathsf{P}\{X < x\}$ is the theoretical distribution function [33]. It can be shown, for an arbitrary continuous distribution function $F(x)$, that

$$\mathsf{P}\left\{D_n < \frac{\lambda}{\sqrt{n}}\right\}$$

tends for $n \to \infty$ to the limit (see [33])

$$K(\lambda) = 1 - 2\sum_{\nu=1}^{\infty}(-1)^{\nu-1} e^{-2\nu^2\lambda^2}.$$

Values of the function $1 - K(\lambda)$ are shown in Table 12.

The goodness-of-fit of the empirical distribution of pseudo-random numbers with the theoretical uniform distribution over the interval (0, 1) is usually tested with a sample size $n \approx 1000$. The use of samples of larger size is extremely cumbersome, since the generation of the variational series (i.e., the series of differences

Table 12

Table of values of the function $1 - K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$

λ	1-K(λ)	λ	1-K(λ)	λ	1-K(λ)
0.28	0.000001	0.67	0.239582	1.06	0.788860
0.29	0.000004	0.68	0.255780	1.07	0.797636
0.30	0.000009	0.69	0.272189	1.08	0.806128
0.31	0.000021	0.70	0.288765	1.09	0.814342
0.32	0.000046	0.71	0.305471	1.10	0.822282
0.33	0.000091	0.72	0.322265	1.11	0.829950
0.34	0.000171	0.73	0.339113	1.12	0.837356
0.35	0.000303	0.74	0.355981	1.13	0.844502
0.36	0.000511	0.75	0.372833	1.14	0.851394
0.37	0.000826	0.76	0.389640	1.15	0.858038
0.38	0.001285	0.77	0.406372	1.16	0.864442
0.39	0.001929	0.78	0.423002	1.17	0.870612
0.40	0.002808	0.79	0.439505	1.18	0.876548
0.41	0.003972	0.80	0.455857	1.19	0.882258
0.42	0.005476	0.81	0.472041	1.20	0.887750
0.43	0.007377	0.82	0.488030	1.21	0.893030
0.44	0.009730	0.83	0.503808	1.22	0.898104
0.45	0.012590	0.84	0.519366	1.23	0.902972
0.46	0.016005	0.85	0.534682	1.24	0.907648
0.47	0.020022	0.86	0.549744	1.25	0.912132
0.48	0.024682	0.87	0.564546	1.26	0.916432
0.49	0.030017	0.88	0.579070	1.27	0.920556
0.50	0.036055	0.89	0.593316	1.28	0.924505
0.51	0.042814	0.90	0.607270	1.29	0.928288
0.52	0.050306	0.91	0.620928	1.30	0.931908
0.53	0.058534	0.92	0.634286	1.31	0.935370
0.54	0.067497	0.93	0.647338	1.32	0.938682
0.55	0.077183	0.94	0.660082	1.33	0.941848
0.56	0.087577	0.95	0.672516	1.34	0.944872
0.57	0.098656	0.96	0.684636	1.35	0.947756
0.58	0.110395	0.97	0.696444	1.36	0.950512
0.59	0.122760	0.98	0.707940	1.37	0.953142
0.60	0.135718	0.99	0.719126	1.38	0.955650
0.61	0.149229	1.00	0.730000	1.39	0.958040
0.62	0.163225	1.01	0.740566	1.40	0.960318
0.63	0.177753	1.02	0.750826	1.41	0.962486
0.64	0.192677	1.03	0.760780	1.42	0.964552
0.65	0.207987	1.04	0.770434	1.43	0.966516
0.66	0.223636	1.05	0.779794	1.44	0.968382

Table 12 (continued)

λ	1-K(λ)	λ	1-K(λ)	λ	1-K(λ)
1.45	0.970158	1.82	0.997346	2.19	0.999864
1.46	0.971846	1.83	0.997533	2.20	0.999874
1.47	0.973448	1.84	0.997707	2.21	0.999886
1.48	0.974970	1.85	0.997870	2.22	0.999896
1.49	0.976412	1.86	0.998023	2.23	0.999904
1.50	0.977782	1.87	0.998145	2.24	0.999912
1.51	0.979080	1.88	0.998297	2.25	0.999920
1.52	0.980310	1.89	0.998421	2.26	0.999926
1.53	0.981476	1.90	0.998536	2.27	0.999934
1.54	0.982578	1.91	0.998644	2.28	0.999940
1.55	0.983622	1.92	0.998744	2.29	0.999944
1.56	0.984610	1.93	0.998837	2.30	0.999949
1.57	0.985544	1.94	0.998924	2.31	0.999954
1.58	0.986426	1.95	0.999004	2.32	0.999958
1.59	0.987260	1.96	0.999079	2.33	0.999962
1.60	0.988048	1.97	0.999149	2.34	0.999965
1.61	0.988791	1.98	0.999213	2.35	0.999968
1.62	0.989492	1.99	0.999273	2.36	0.999970
1.63	0.990154	2.00	0.999329	2.37	0.999973
1.64	0.990777	2.01	0.999380	2.38	0.999976
1.65	0.931364	2.02	0.999428	2.39	0.999978
1.66	0.991917	2.03	0.999474	2.40	0.999980
1.67	0.992438	2.04	0.999516	2.41	0.999982
1.68	0.992928	2.05	0.999562	2.42	0.999984
1.69	0.993389	2.06	0.999588	2.43	0.999986
1.70	0.993828	2.07	0.999620	2.44	0.999987
1.71	0.994230	2.08	0.999650	2.45	0.999988
1.72	0.994612	2.09	0.999680	2.46	0.999989
1.73	0.994972	2.10	0.999705	2.47	0.999990
1.74	0.995309	2.11	0.999723	2.48	0.999991
1.75	0.995625	2.12	0.999750	2.49	0.9999920
1.76	0.995922	2.13	0.999770	2.50	0.9999925
1.77	0.996200	2.14	0.999790	2.55	0.9999956
1.78	0.996460	2.15	0.999806	2.60	0.9999974
1.79	0.996704	2.16	0.999822	2.65	0.9999984
1.80	0.996932	2.17	0.999838	2.70	0.9999990
1.81	0.997146	2.18	0.999852	2.75	0.9999994

$|F_n(x) - F(x)|$) for $n = 1000$ is a very difficult operation even when high-speed computers are used. For example, the generation of the variational series and the evaluation of the quantity

$n^{1/2}D_n$ were carried out on the "Strela" computer, the total time taken for testing by Kolmogorov's criterion being 15-20 minutes. The hypothesis of "uniformity" was rejected if the quantity $n^{1/2}D_n$ exceeded 1.5. The result $n^{1/2}D_n<0.5$ is also very undesirable. The existence of an "extremely good" fit would indicate that the "randomness" of the pseudo-random numbers is questionable.

The interval (0.7, 1.0) is conveniently taken as the optimum range of values of the quantity $n^{1/2}D_n$.

Together with the χ^2 and Kolmogorov criteria, the ω^2 criterion is also often employed. Its advantage over the χ^2 criterion is the fact that it is based only on the individual values of the sample and not on data based on subdivision into intervals.

The basis of the ω^2 criterion is the statistic

$$\omega^2=\int_{-\infty}^{+\infty}[F_n(x)-F(x)]^2\,dF(x),$$

where $F(x)$ is the theoretical and $F_n(x)$ is the empirical distribution function (n is the sample size). If we arrange the sampling values $x_1, x_2, \ldots, x_n$ in a variational series, we have, in the case of a continuous function $F(x)$,

$$\omega^2=\frac{1}{12n^2}+\frac{1}{n}\sum_{\nu=1}^{n}\left[F(x_\nu)-\frac{2\nu-1}{2n}\right]^2.$$

The distribution of the random quantity ω^2 does not depend on $F(x)$.

Since the Kolmogorov and ω^2 criteria require the construction of a variational series, high-speed computers are conveniently used for evaluating the statistics $n^{1/2}D_n$ and ω^2 whenever the sample size n is fairly large. According to the ω^2 criterion ($n=1000$), the hypothesis of "uniformity" must be rejected if $\omega^2>0.0005$.

Table 13 shows the results of tests of pseudo-random numbers using at the same time the three criteria: the χ^2, the Kolmogorov and the ω^2 criteria. The pseudo-random numbers tested were distributed uniformly over the interval (0, 1) and were generated on the "Strela" computer by means of a three-instruction program. Tests were carried out for 10 different initial values of α_0. The size n of the sample tested was 10,000.

It can be seen from the table that for all α_0 we have $10<\chi^2<30$ and there are thus no grounds for rejecting the hypothesis.

Table 13

Serial number of values of α_0	χ^2	$D_n\sqrt{n}$	ω^2
1	15.4	0.84	0.00039
2	14.6	0.69	0.00013
3	20.1	1.10	0.00027
4	17.2	0.74	0.00009
5	17.4	0.62	0.00019
6	24.1	1.04	0.00014
7	12.8	0.91	0.00007
8	19.4	0.87	0.00015
9	16.5	0.79	0.00044
10	26.7	0.65	0.00011

It can be noticed, in addition, that the distribution of ten sample values of χ^2 follows in general the χ^2 distribution with 19 degrees of freedom.

Thus, with this distribution, the majority (about 50%) of the values of the quantity must occur in the interval (15, 25), and this is in complete agreement with the experimental data. This may serve as a basis for accepting the hypothesis of uniformity of the distribution of pseudo-random numbers.

The distribution of the value of $n^{1/2} D_n$ as shown in the table agrees with the limiting distribution of Kolmogorov's criterion. The majority of values occurs in the optimum interval (0.7, 1.0), which corresponds to the theoretical calculations. The calculations using Kolmogorov's criterion confirm the validity of the hypothesis.

In checking the "randomness" and the distribution of pseudo-random numbers it is expedient to check the correctness of the operation of the program as a whole. This can be done by evaluating, in a control problem, some parameter or other the value of which is known a priori. Of course, the calculation of the value of the parameter is carried out by the Monte Carlo method using pseudo-random numbers. If the calculated result does not depart from the a priori value more than the theoretically calculated statistical error, this means that the defects of the pseudo-random numbers do not affect the behavior of the calculations and can be accepted. In the opposite case we must modify the program and repeat the control computation until a positive result is obtained. We can choose as the control problem the calculation of the volume of an n-dimensional hypersphere. As is well known, the volume of an n-dimensional hypersphere is equal to

$$V_n = \frac{2}{n} \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}.$$

If E_n is the hypercube $-1 \leqslant x_i \leqslant +1$ $(1 \leqslant i \leqslant n)$, the ratio of the volumes V_n and E_n

$$v_n = \frac{V_n}{E_n} = \frac{\pi^{n/2}}{n \cdot 2^{n-1} \Gamma\left(\frac{n}{2}\right)}$$

may be calculated by the Monte Carlo method.

In concluding, we observe that satisfactory results, obtained by using only once some goodness-of-fit criterion for testing pseudo-random numbers, do not guarantee the quality of these numbers. Only a repeated investigation of pseudo-random numbers by means of some goodness-of-fit criterion or other, based on comparing the theoretical distribution with the empirical distribution of the statistic on which the criterion is based (i.e., χ^2, $n^{1/2} D_n$ or ω^2), can justify us in assuming that a given method of obtaining sequences of pseudo-random numbers is fully satisfactory.

4. PHYSICAL GENERATORS OF UNIFORM RANDOM NUMBERS

Random numbers uniformly distributed over the interval (0, 1) can be obtained by means of a random physical process. To do this, special "attachments" to electronic computers, the so-called random-number generators, have been developed and constructed. The use of random-number generators on an electronic computer enables us to obtain a combination of random binary digits in the array of digits of the random number represented in the electronic computer.

There are two main methods of obtaining random numbers by means of a physical process. The first method is based on the radiation of radioactive substances, the second on the noise of electron tubes. According to the physical process used, random-number generators are subdivided into radioactive and radio-noise type.

Let us proceed to consider radio-noise type random-number generators. Firstly we shall describe the noise sources most often used in such generators, after which we will pass to describe circuits of the random-number generators themselves.

The basis of the operation of any radio-noise type random-number generators is the noise source—a noise generator. In a whole series of electronic circuits there is inherent fluctuating noise, which, with suitable amplification, can ensure sufficient fluctuation of the output voltage. Such internal sources of fluctuations comprise:

a) thermal noise, caused by the motion of electrons in the space-charge volume of a tube;

b) shot noise, caused by departure from uniformity (in time) of the number of electrons leaving the cathode surface; and a whole series of other sources.

External sources of fluctuations comprise variable electronic and magnetic fields of electric generators, cosmic radio noise, gas-discharge devices, etc. All these sources of fluctuations can be used for designing noise generators.

Several typical noise-generator circuits which give good results in random-number generators have been developed. The noise-generator circuits shown below have been suggested by V. P. Smiryagin at the Computing Center of the USSR Academy of Sciences [26, 28].

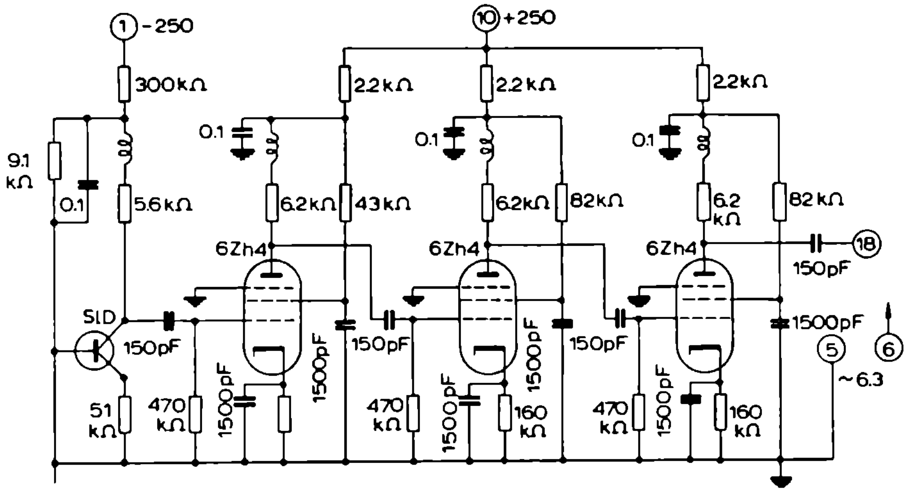


Fig. 30. Noise generator with germanium transistor.

Figure 30 shows a noise-generator circuit using an S1D germanium transistor as the noise element. The noise generator produces a series of random numbers of pulses which are amplified and shaped into standard pulses. Subsequent transformations in the random-number generator develop from these pulses the binary code representation of a random number.

Figure 31 shows a noise-generator circuit the noise element of which uses a gas-discharge tube and a magnet. Noise pulses of the order of several tens of millivolts can be obtained directly from the gas-discharge tube for a suitable orientation of the magnet. The output signals from the tube are applied to the input of an amplifier, and working signals with amplitude of 25 to 40 volts are obtained at the output of the circuit.

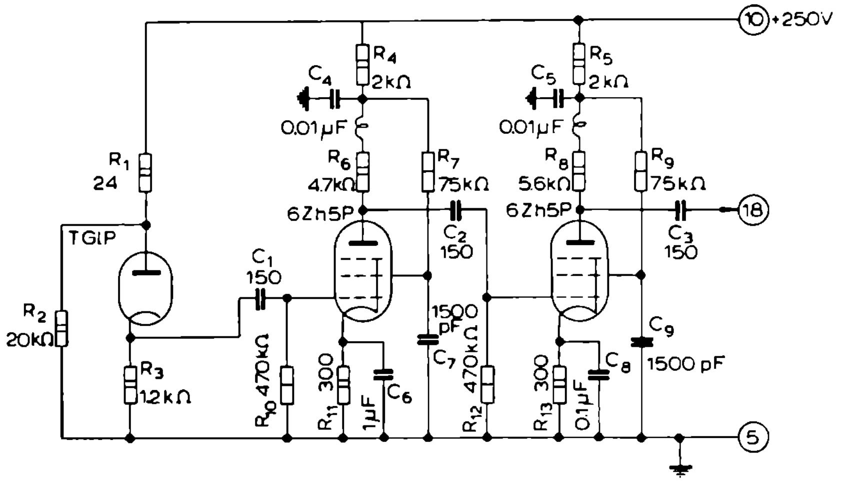


Fig. 31. Noise generator using gas-discharge tube with magnet.

Among the noise-generator circuits considered, the circuits using a gas-discharge tube with a magnet give the best results for the operation of random-number generators.

Of the noise generators produced commercially we must recommend the GShN noise generator, which covers a frequency band from 50 cps to 6 Mc with an effective value of the output signals of 0.75 volt and an output resistance of 75 ohms + 1%.

Radio-Noise Type Random-Number Generator Circuits

There are a variety of configurations of the circuits of random-number generators which can be classified accordingly into three groups:

1. Random-number generator circuits which make use of the random state of the system (of the circuit) after perturbation (after switching on, switching off, start, stop).
2. Random-number generator circuits in which time intervals of random duration are registered. During these various (random) intervals of generation, standard pulses of a fixed frequency are sent to a counter or a trigger.
3. Random-number generator circuits in which the number of random pulses during a fixed interval of time Δt is registered.

We give a few examples of random-number generators corresponding to the various groups enumerated above.

1. One of the typical random-number generators working according to the principle of the random state of a system is that developed by V. V. Chavchanidze [64]. This random-number generator is electromechanical and is a small electric motor having a speed of rotation of about 3000 revolutions per minute.

In order to remove resonance oscillations, the motor is mounted on a massive base. A circular disk of 10 cm diameter is fixed on the shaft of the motor. The circumference of the disk is divided in 100 equal parts numbered from 0 to 99. A fixed pointer is mounted against the disk. The motor is connected for 2-3 seconds, is then disconnected and is brought to a stop by means of a brake. As the disk reaches its position of rest the pointer indicates a certain division on it. The corresponding random number (division) is recorded in a table and the motor is again connected for obtaining the following random number.

In view of the fact that from considerations of symmetry the pointer can be found with equal probability against any one of the divisions (from 0 to 99) such a random-number generator produces uniformly distributed random numbers assuming discrete values from 0 to 99. The generation of a single random number in the electromechanical random-number generator takes a few seconds. In comparison with the operation of electronic computers this process of generating random numbers is very slow. The low speed and the manual operations involved (start, stop, recording of disk divisions) make such a type of random-number generator unacceptable for electronic computers. However, it can be usefully employed when there is a need for a comparatively large quantity of random numbers without the use of electronic computers.

2. Another interesting random-number generator scheme based on the principle of the random state of a system is that developed by Z. Pawlak (Poland) [228]. Pawlak's random-number generator uses the random state of an electronic circuit, rather than that of an electromechanical one as in the previous example. A circuit diagram taken from [228] of the Pawlak flip-flop trigger circuit is shown in Fig. 32. The key K serves to connect the anode voltage.

Each switching of the key K sets the trigger (the random system) in one of two possible states: A or B. We can obtain, by means of this circuit, $2k$ random elements X_j, where $X_j = A$ if the jth switching of the key K sets the trigger in the position A, and $X_j = B$ if the jth switching of the key K sets the trigger in the position B.

Thus a finite set of statistically independent results A and B can be obtained. We give, below, one of the sets of positions of the trigger after switching: $AABAABBABBAABBBABBAB$.

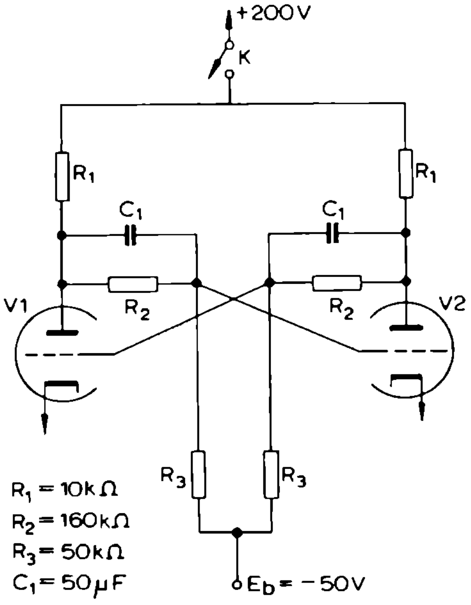


Fig. 32. Circuit diagram of trigger.

Let γ_k be a set of k pairs of elements X_{2k} such that $\gamma_i = x_{2i-1}$, x_{2i}, where $1 \leqslant i \leqslant k$. By omitting in the sequence γ_k all elements AA and BB, we obtain a third set the elements of which are only pairs of the type AB and BA. Let us denote these pairs 0 and 1 respectively.

Let $P_j(A)$ and $P_j(B)$ be the probabilities, for the jth switching of the contact K, that the trigger is set in the positions A and B respectively.

By assuming the trigger to be asymmetrical

$$P_j(A) > P_j(B),$$

and the properties of the trigger to be constant between two switchings, we can write

$$P_{2i-1}(A) = P_{2i}(A), \tag{6.1}$$

$$P_{2i-1}(B) = P_{2i}(B). \tag{6.2}$$

From (6.1) and (6.2) we have

$$P_{2i-1}(A)\,P_{2i}(B) = P_{2i}(A)\,P_{2i-1}(B). \tag{6.3}$$

Since

$$P_{2l-1}(A)\,P_{2l}(B)=P_l(0) \tag{6.4}$$

and

$$P_{2l-1}(B)\,P_{2l}(A)=P_l(1), \tag{6.5}$$

then

$$P_l(0)=P_l(1), \tag{6.6}$$

where $P_i(0)$ and $P_i(1)$ are the probabilities of obtaining the values 0 and 1 respectively at the ith place of the sequence γ_k. As has been indicated above, by zero (0) and one (1) are denoted pairs of successive states of the trigger of the type AB and BA respectively. Such a trigger circuit enables us to obtain a random state of a single binary digit. By means of N such switching triggers a generator of N-binary-digit random numbers can be assembled.

3. A random-number generator prototype using Pawlak's circuit has been developed at the Computing Center of the USSR Academy of Sciences for the BESM-II computer. This generator belongs to the second group (based on the principle of a random interval of generation).

Figure 33 shows the block diagram of this random-number generator. Noise signals are generated by a noise generator NG consisting of a TG1P gas-discharge tube and a pulse-shaper output tube. Pulses from the noise generator are applied, in a correction stage, to a trigger T_1. The noise pulses flip the triggers T_1 and T_2 in random sequence, and in the intervals between them pulses of the fixed frequency $f=200$ kc can pass. The trigger T_1 can have two possible states, A or B. When the random-number generator operates, a random set of positions of the trigger T_1 of the following possible form is obtained:

$$ABBABBBAABAAA.$$

From the random set of positions of the trigger the special correction circuit removes elements of the form AA and BB and, by means of the second trigger T_2, generates a new sequence the elements of which are only pairs of the type AB and BA. We denote these pairs 0 and 1 respectively.

It can be shown by similar arguments to the above that

$$P_i(0)=P_i(1).$$

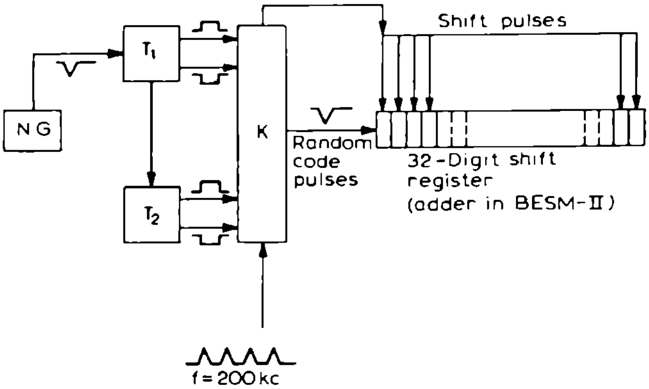


Fig. 33. Simplified block diagram of random-number generator for the BESM-II.

In other words, the probability that the ith pulse sets the trigger T_1 in the "zero" state is equal to the probability that it sets the trigger in the "one" state.

The above correction circuit thus enables us to obtain equiprobable random binary digits. The correction stage generates random code pulses (which arrive at the first digit of the shift register), as well as shift pulses which, via the common bus, reach all 32 digits of the adder.

4. Random-number generators based on recording a random number of pulses in a fixed interval of time are used today in several scientific organizations of the Soviet Union. We give a general description of the random-number generator designed for the "Strela" computer at the Computing Center of the USSR Academy of Sciences. A detailed description of the circuit and mode of operation of the generator is given in [26].

The noise-signal source is a noise generator comprising a TG1P gas-discharge tube with a magnet (see the circuit in Fig. 31). The block diagram of the random-number generator is shown in Fig. 34. A front view of the random-number generator for the "Strela" computer is shown in Fig. 35. The random-number generator is intended for obtaining simultaneously all the digits of twelve-digit binary numbers. All twelve digits work independently of the others from individual noise generators. For each digit of the generator there are a noise generator, a gate, a pulse shaper, a trigger and an output inverter.

At each step of operation of the "Strela" computer the random-number generator produces a random number which reaches a fixed location of the accumulator of constants (the location has the ordinal number 7757). The transfer of the random number from the location 7757 to the working location a is carried out by means of the instruction $\boxed{7757}\boxed{-}\boxed{a}\boxed{0}\boxed{13}$.

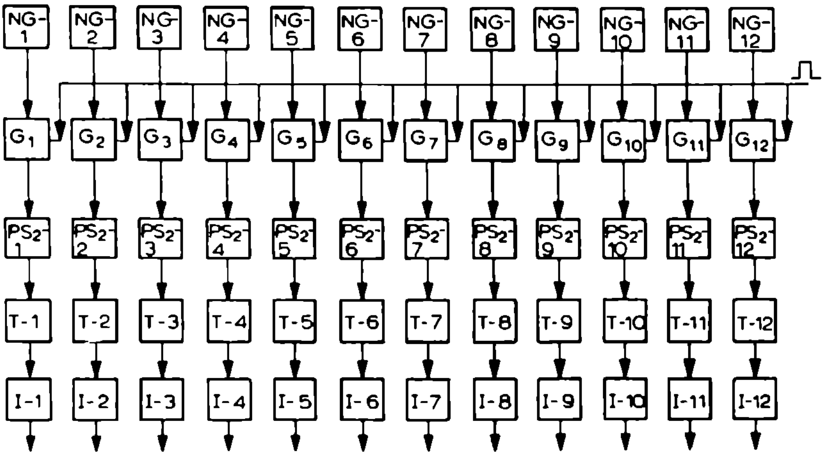


Fig. 34. Block-diagram of random-number generator for the "Strela" computer.

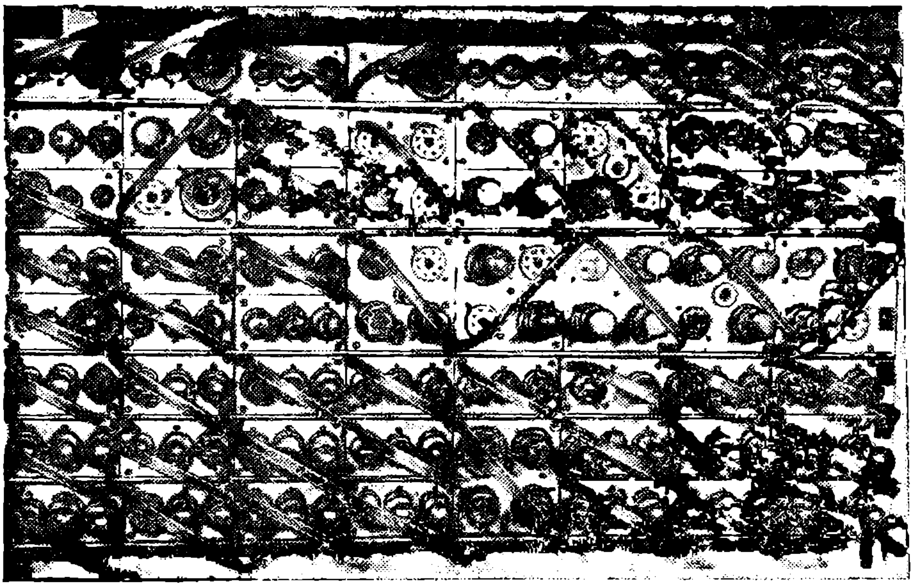


Fig. 35. Front view of random-number generator for the "Strela" computer.

Let us proceed to describe the operation of the random-number generator. A random number of pulses from the noise generator is routed via the gate and the pulse shaper to the count input of the trigger. The circuit diagram of the gate is shown in Fig. 36. It has two control inputs. To the first input there are applied pulses from the noise generator, and to the second input there are applied positive gating pulses from a one-shot multivibrator. In order to ensure that random numbers be issued at each step of operation of the computer, the one-shot multivibrator is triggered at the beginning of a step by a special pulse applied from the control unit of the "Strela" computer. Pulses occur at the gate output for the coincidence of a wide gating signal from the one-shot multivibrator with pulses from the noise generator.

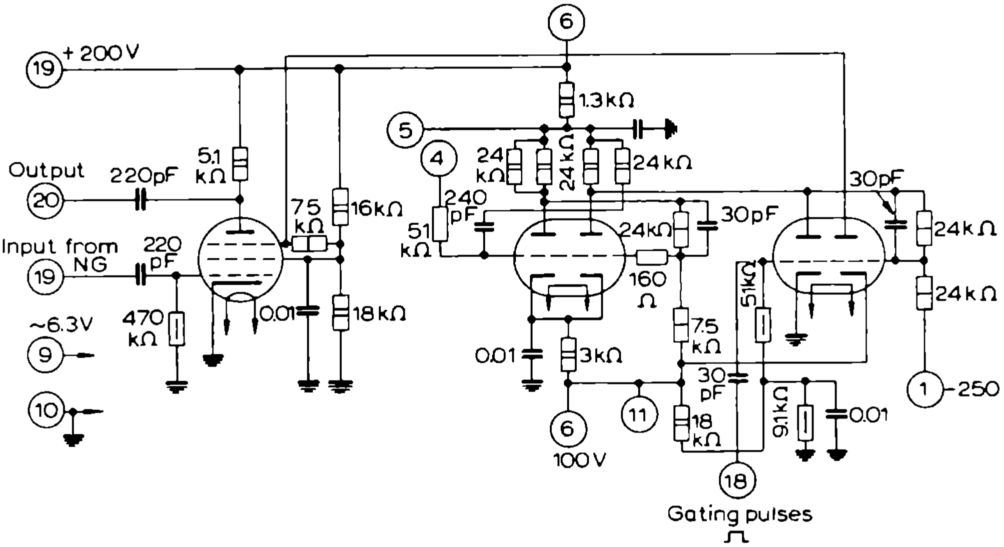


Fig. 36. Circuit diagram of gate of random-number generator.

Pulses from the gate are applied, via the pulse shaper, to the count input of the trigger, the circuit diagram of which is shown in Fig. 37. The trigger records the random number of pulses applied at its input (even or odd) by assuming, when a train of pulses has ended, one of two random states: zero (0) or one (1). The random states of all twelve triggers (0 or 1) are transmitted, via the inverters, to the corresponding digits of the 7757 location of the accumulator of constants. The circuit diagram of the inverter is shown in Fig. 38. The twelve digits of the random-number generator are used as the last twelve digits of the mantissa of the number.

As regards construction, all stages of the random-number generator are standard removable units.

F. Sterzer (England) has constructed a random-number generator using subharmonic oscillators [245]. The random-number generator is a continuous-action electronic device capable of generating random binary symbols at a rate of the order of 3×10^7 symbols per second. The generating rate can be increased by connecting several such devices in parallel.

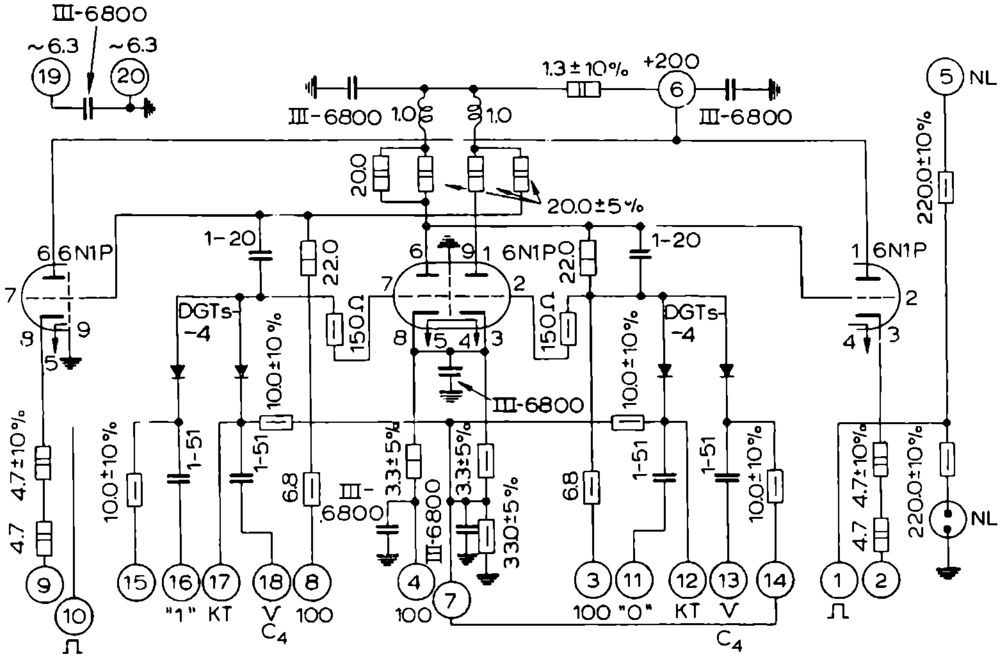


Fig. 37. Circuit diagram of trigger of random-number generator.

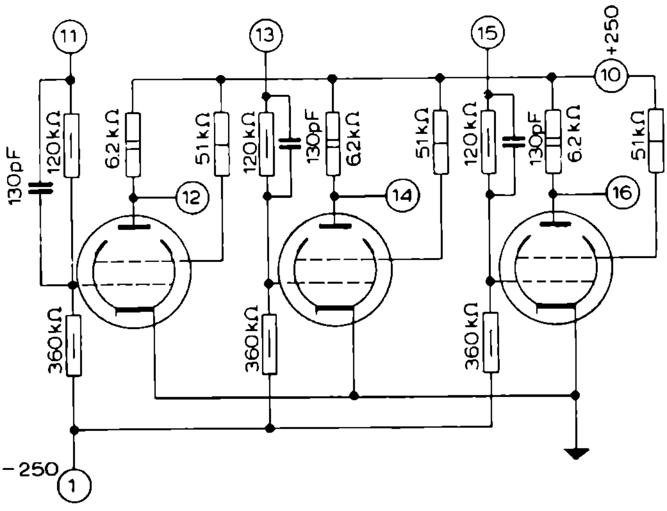


Fig. 38. Circuit diagram of inverter of random-number generator.

The principle of operation of the equipment consists in the following. Two subharmonic oscillators are driven by a common source of oscillations of a frequency of about 4000 Mc. The frequency of the output signals of the oscillators is halved to about 2000 Mc. Owing to the presence of noise the phases of the output signals are quantities random in time. The random device compares the phases of the output signals of the generators and generates a sequence of binary signals. Ones occur in the case of coincidence of phases, and zeros if the signals are found in antiphase. The equipment comprises a klystron oscillator, waveguides and crystal diodes.

We should mention also the random-number generator developed by I. A. Danil'chenko. The mathematical bases of this attachment for the "Strela" electronic computer are described in [30]. The key idea consists in balancing the probabilities of occurrence of binary figures in the digit positions of the random-number generator by superimposing direct and inverse representations. The random-number generator obtains random binary digits with close to 0.5 probability of occurrence of a zero, and is characterized by stable operation.

Let us proceed to describe radioactive random-number generators. These usually consist of a radiation source of radioactive particles and a counter. The latter counts radioactive particles registered during a certain interval of time Δt. If this number is even, the random digit of the generator connected with the corresponding counter is taken equal to zero. If, however, the number of particles is odd, the value of the random digit is taken equal to one. We can easily calculate that if the probability of radiation of k radioactive particles by the radioactive substance during the time Δt is equal to

$$W_k = \frac{(\lambda\,\Delta t)^k\, e^{-\lambda\,\Delta t}}{k!},$$

then the probability of occurrence of an even number of particles during the time Δt is equal to

$$p = \frac{1 + e^{-2\lambda\,\Delta t}}{2}.$$

If $\lambda\,\Delta t$ is sufficiently large, the value of p is close to 0.5.

A good radiation source of radioactive particles is radioactive cobalt. Figure 39 shows the circuit diagram of the counter of a radioactive random-number generator operating with β particles radiated by radioactive cobalt. An STS-2 type β-particle counter is connected to the input of a three-stage amplifier. This counter is shown in Fig. 40. The use of transmitters of particles safe for the human operator has given comparatively low frequencies. For example the luminous dial-plate of watches has given random pulses with a repetition frequency of a few cycles per second. More intense sources of β-particles enable us to obtain good results in the operation of a random-number generator.

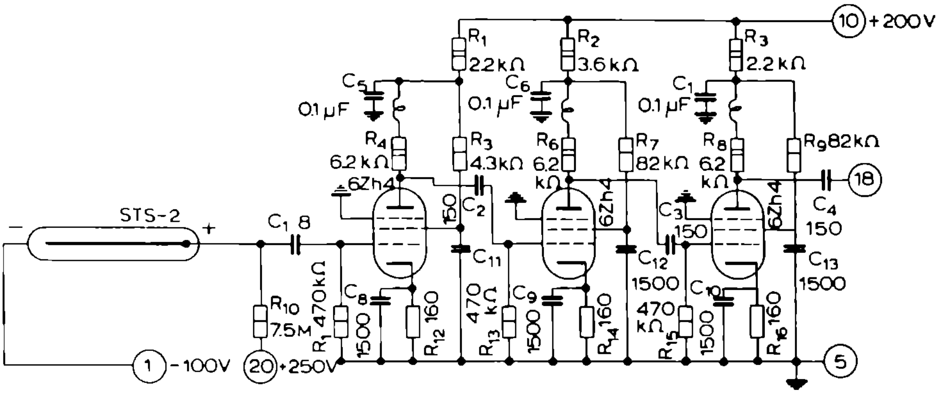


Fig. 39. Circuit diagram of counter of radioactive type random-number generator.

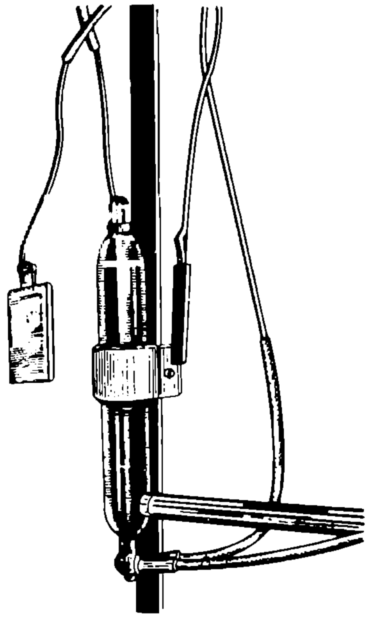


Fig. 40. Counter of radioactive type random-number generator with slabs of radioactive cobalt.

A generator using a counter of radioactive radiation has been installed on the G-2 electronic computer at Göttingen at the Max Planck Physical Institute. The block diagram of this generator is shown in Fig. 41. It uses a single-digit counter of radioactive radiation of γ-quanta. Random states "0" and "1" are applied from the trigger to the input of the gate. The trigger is flipped after each registration of a γ-quantum. The generator gives 800 random symbols per second for a number of recorded particles up to 4000 γ-quanta per second.

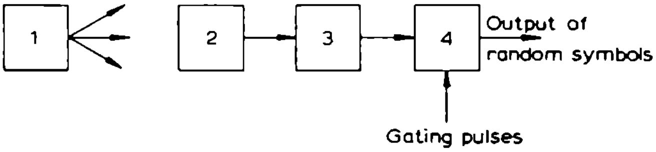


Fig. 41. Random-number generator of the G-2 computer at Göttingen: 1—radioactive preparation, 2—counter tube, 3—trigger, 4—gate.

The purpose of a random-number generator is to obtain random numbers uniformly distributed over the interval (0, 1). Therefore, a test of the quality of the numbers obtained is at the same time also a check of the operation of the generator itself. The need of periodically assessing the random numbers generated by means of a random-number generator makes it necessary to establish criteria of its correct operation and to develop for it suitable testing procedures.

One of the most important characteristics of a random-number generator is the estimate of the probability of occurrence of a zero or one in each digit of a random number or, in other words, the estimate of the maximum departure of the probability of occurrence of a zero or one from its prescribed value 0.5. A value of 0.5 for this probability will occur only in the case of an ideal random-number generator giving an exact rigorously uniform distribution. In practice, however, such a situation never occurs. Lack of constancy of parameters of the circuit, reduced emission of tubes and other reasons bring about the mismatching of the random-number generator, and hence departure from uniformity of the distribution of the random numbers generated by it. We assume that the probability of occurrence of a zero in each digit is not equal to 0.5 but to $0.5(1+\delta)$, and the probability of occurrence of a one is accordingly $0.5(1-\delta)$.

There are several effective methods of decreasing the value of δ. Thus good results are obtained, for example, by replacing the sequence of random binary numbers of the random-number generator $\{\xi_i\}$ by another random sequence $\{\eta_i\}$, where $\eta_i = \xi_{i-1} + \xi_i \pmod 2$. Thus the value of η_i is one if the value of ξ_{i-1} and ξ_i are different, and is zero in the opposite case. Thus

$$\mathsf{P}\{\eta_l=0\}=\mathsf{P}\{\xi_{l-1}=0\}\,\mathsf{P}\{\xi_l=0\}+\mathsf{P}\{\xi_{l-1}=1\}\,\mathsf{P}\{\xi_l=1\}= \frac{1}{4}(1+\delta)^2+\frac{1}{4}(1-\delta)^2=0.5\,(1+\delta^2).$$

Similarly $\mathsf{P}\{\eta_l=1\}$ is equal to $0.5\ (1-\delta^2)$. Thus the departure obtained is of a higher order of smallness (δ^2 instead of δ) (see [30]).

Such a method can be realized in practice by adding random digits modulo 2 in the counter stage of a random-number generator.

Let us analyze in greater detail some characteristics of the operation of random-number generators. Suppose a sequence of independent random binary numbers $\{\xi_l\}$ obtained by means of a random-number generator has a nonuniform distribution of zeros and ones. Let the probability of occurrence of a zero for an element ξ_l be equal to $0.5\ (1+\delta)=p$ and the probability of a one be accordingly $0.5\ (1-\delta)$. The length n of this sequence is assumed to be sufficiently large.

One of the most important statistical characteristics of a sequence $\{\xi_l\}$ is the estimate of the number of zeros (or ones) for sufficiently large n.

In particular, it is of interest to estimate the relative departure of $P(n,\ m,\ p)$ from $P(n,\ m,\ 0.5)$ for large n where

$$P(n,\ m,\ p)=\binom{n}{m}p^m(1-p)^{n-m}$$

is the probability of occurrence of m zeros (or ones) in an n-symbol sequence of binary numbers.

By using the de Moivre-Laplace theorem, it can be shown that

$$\Delta=\frac{P(n,\ m,\ p)-P(n,\ m,\ 0.5)}{P(n,\ m,\ 0.5)}\sim\delta\sqrt{n}\,C,$$

where C is a constant which for sufficiently large n can be made arbitrarily small. The relative departure Δ is directly proportional to δ and $n^{1/2}$.

It is of interest to fit this result with the corresponding estimate in the case of small values of n. We point out that all upper estimates given below refer to the absolute value of the departure Δ.

We assume, as before, that the probability of occurrence of a zero in each digit of the random-number generator is equal to $p=0.5(1+\delta)$, and that the probability of a one is accordingly $q=0.5(1-\delta)$. Suppose the random-number generator produces random n-digit numbers. We shall estimate the relative departure of the probability $P(n,\ m,\ p)$, that a random number contains m zeros, from the nominal probability $P(n,\ m,\ 0.5)$.

By using Bernoulli's theorem we obtain

$$|\Delta| \leqslant (1+|\delta|)^n - 1.$$

If it is required to find a δ such that $|\Delta| < \Delta_0$ ($1 > \Delta_0 > 0$ is an assigned number), then the corresponding $|\delta_0|$ must be chosen equal to $(1+\Delta_0)^{\frac{1}{n}} - 1$.

In fact by putting $(1+|\delta_0|)^n - 1$ equal to Δ_0 we obtain the required result.

Thus $|\Delta| < \Delta_0$ if $|\delta| < |\delta_0|$.

We observe that for small values of δ there exists for $|\Delta|$ a simple linear upper estimate.

Let $0 < \delta < \frac{1}{n^2}$. Then

$$\Delta < \delta(n+0.5).$$

Thus for small values of n an upper estimate for the relative departure Δ is proportional to n, and not to $n^{1/2}$, as in the case of large values of n. In fact the departure Δ is of the same order of smallness as δn only for values of m that are found at the ends of the interval $(0, n)$. For $m \sim \frac{n}{2}$ the value of Δ is close to zero.

For large values of n the partial departures Δ, accumulating on each other, give as a result an overall total departure. This total departure lies between the largest and the smallest partial departures and therefore is not proportional to n but only to $n^{1/2}$. Let us estimate the relative departure of the distribution function of the number of zeros (or ones) for the case $p = 0.5(1+\delta)$ from that with $p = 0.5$ for small values of n. We have

$$\mathrm{P}\{\mu \leqslant m\} = I_{1-p}(n-m;\ m+1) = \frac{1}{B(n-m;\ m+1)} \int_0^{1-p} s^{n-m-1}(1-s)^m\, ds.$$

The relative departure Δ will have in this case the following form

$$\Delta = \frac{\int_{\frac{1}{2}-\delta/2}^{1/2} s^{n-m-1}(1-s)^m\, ds}{\int_0^{1/2} s^{n-m-1}(1-s)^m\, ds};$$

it can be shown that

$$\Delta < \delta(n - m)$$

or, since $n - m \leqslant n$,

$$\Delta < \delta n.$$

The above estimate for Δ shows that the departure Δ decreases for an increase of m and is equal to zero for $m = n$.

In order that the value of the quantity δ be sufficiently small in each digit of the random-number generator, a series of design requirements must be imposed on the individual component parts of the random-number generator. Of course, such requirements must be theoretically justified.

Let us consider the simplest version of random-number generator design (a noise generator, a gate, a trigger and an inverter) and let us find the dependence of the quantity δ on various parameters of the trigger. It is well known that asymmetry of the latter plays an important role in the nonuniformity of the corresponding random digit.

The noise generator in a random-number generator consists of a noisy stage and an amplifier. From the amplifier the pulses are applied, via a gate and a pulse shaper, to the trigger of the random-number generator.

As is well known, a trigger can assume two opposite states: a state A (for which the counter stage assumes the value zero) and a state B (for which the corresponding value is one).

In order to flip the trigger from one state to the other, a pulse of a certain amplitude must be applied to its count input. Suppose the minimum pulse amplitude capable of flipping the trigger from state A to state B is equal to a_1. At the same time a pulse that flips the trigger from one state to another may be insufficient to reverse the state.

Without loss of generality we can assume that the minimum pulse amplitude capable of accomplishing the reverse flipping of the trigger from state B to position A is equal to a_2 and that $a_2 > a_1$. Thus a pulse having amplitude larger than a_2 accomplishes the flipping of the trigger in either direction.

Pulses of even larger amplitude are characterized by the property of accomplishing a double flipping of the trigger: from one state to another and back. Let the minimum amplitude of such a pulse be equal to $a_3 > a_2$.

Let p_1 be the probability for a pulse to have amplitude less than a_1:

$$\mathsf{P}\{u < a_1\} = p_1.$$

Let p_2 be the probability for the pulse amplitude to lie between a_1 and a_2:

$$\mathsf{P}\{a_1 \leqslant u < a_2\} = p_2.$$

Let p_3 be the probability for the pulse amplitude to lie between a_2 and a_3:

$$\mathsf{P}\{a_2 \leqslant u < a_3\} = p_3.$$

Finally let p_4 be the probability that the pulse amplitude is not smaller than a_3:

$$\mathsf{P}\{u \geqslant a_3\} = p_4.$$

We denote by P_n the probability that after the arrival of the nth pulse the trigger assumes the state A, and by Q_n the probability of the opposite event.

We denote by P_0 the probability that, before the random-number generator starts to operate, the trigger is found in the state A, and by Q_0 the probability of the opposite event.

Of course, we have to estimate the probability of occurrence of zero (or one) in a digit of the random-number generator for the stationary (steady-state) process of obtaining random binary digits from the corresponding trigger. To do this we must study the behavior of P_n in the transient process that leads to stationary operating conditions of the random-number generator.

By applying the elementary theory of Markov processes, we can prove the following:

The probability of nonoperation of the trigger after a pulse has arrived from the noise generator is p_1, and p_2, p_3 and p_4 are the probabilities of unilateral, bilateral and double flipping of the trigger. Then, in the process of establishing stationary operating conditions, the probability of occurrence of a zero at the corresponding digit is expressed by the formula

$$P_n = \frac{p_3}{p_2 + 2p_3} + \left(P_0 - \frac{p_3}{p_2 + 2p_3}\right)(1 - p_2 - 2p_3)^n,$$

where, as stated above, P_n is the probability that the trigger assumes the zero state after the arrival of the nth pulse and P_0 is the probability of the same event before the generator starts to operate.

For sufficiently large n, P_n can be taken equal to

$$P = \frac{p_3}{p_2 + 2p_3}.$$

Hence

$$Q = 1 - P = \frac{p_2 + p_3}{p_2 + 2p_3},$$

$$Q - P = \frac{p_2}{p_2 + 2p_3}.$$

If $p_2 \ll p_3$, then

$$Q - P \sim \frac{1}{2} \frac{p_2}{p_3}.$$

If we consider all pulses to be operative, then, by putting $p_1 = 0$, we obtain

$$Q - P = \frac{p_2}{p_2 + 2p_3} = \frac{p_2}{1 + p_3 - p_4}.$$

If $p_2 \ll p_3$ and $p_4 \ll p_3$, then $p_3 \sim 1$ and therefore

$$Q - P \sim \frac{1}{2} p_2.$$

If we put the difference $Q - P$ equal to 0.001, then $p_2 \sim 0.002$.

Since $Q - P = \delta$, then

$$\delta \sim \frac{1}{2} p_2.$$

If the voltage of the noise generator is equal to V_{NG} volts, then, applying the formula $p_2 \approx 2\delta$, the spread of the operating voltages (reduced to the amplifier input) must be not larger than

$$V_S = 2\delta V_{NG}.$$

If the gain of the amplifier is equal to k, then the spread of the operating levels of the trigger must be not larger than

$$\Delta V = kV_S = 2k\delta V_{NG}.$$

If $k = 40$, $\Delta = 0.01$, $V_{NG} = 2.5$ v and $n = 12$, then $\Delta V \approx 0.2$ v.

If the mean operating level of the trigger is equal, for example, to 15 v, then we must provide in the trigger such a stability of the resistances and voltages that the relative spread of the operating levels shall not exceed $\frac{0.2}{15} \cdot 100\% \approx 1.3\%$.

Such requirements, arrived at theoretically, set very stringent limitations on the parameters of random-number generators. Moreover they are not met in the great majority of random-number generators satisfactorily operating today. However, they can often be dispensed with, since the quality of random numbers generated by means of random-number generators is assessed basically by means of the system of accepted statistical criteria which must be satisfied by the aggregate of random numbers.*

Several random-number generators which provide stable and sufficiently high-grade service are now in use. They satisfy a system of statistical tests despite the fact that in individual cases the departures of the probabilities of occurrence of individual combinations of digits exceed critical values.

In order to illustrate the potentialities of random-number generators, we shall cite, for example, data of a random-number generator operating on a "Strela" computer at the Computing Center of the USSR Academy of Sciences [26].

Experience gained in solving a whole series of problems by means of this generator as well as the results of repeated tests enable us to draw the following conclusions:

1) The random-number generator gives a relative error in the results which does not exceed 10-12%, but this error is less than 10% for the great majority of problems and lies within the range 4-10%.

2) The probability of occurrence of a zero or a one at an arbitrary binary digit of a number generated by means of the random-number generator varies within the interval $0.44 < p_i < 0.56$. This estimate has been obtained as a result of a series of tests extending over a long period of time.

3) For the majority of problems solved by means of random-number generators the result differs from the result obtained by means of pseudo-random numbers by less than the evaluated statistical error within a single σ.

5. TESTS OF THE OPERATION OF RANDOM-NUMBER GENERATORS

The quality of random numbers obtained by means of random-number generators must be periodically tested. This is necessary for preventive maintenance of the installation, since the process of random-number generation can go out of control: the generator can "run wild."

---

*In addition, the probabilities in the individual digits can be evened out by several special methods described in (Chapter I, Section 3) and also in [26, 28, 30].

Therefore, on all electronic computers containing random-number generators periodic test routines are run several times a day in order to check the performance of the random-number generator. As a rule, test programs include [26]:

a) A program to test "randomness" by using the method of "runs" (see Section 3). Tests according to this criterion usually give perfectly good results.

b) Tests for checking departure of the empirical distribution from uniform. The goodness-of-fit criteria usually employed are the χ^2 criterion and Kolmogorov's criterion, although other, related methods are often used. Thus at the Computing Center of the USSR Academy of Sciences distribution tests are carried out according to the χ^2 criterion on k groups of 10,000 numbers each ($k \approx 5$—7).

If for any group tested the value of χ_q^2 exceeds 30, then it is assumed that the distribution does not coincide with the uniform distribution regardless of the results of tests of the remaining groups. In this case the degree of confidence of our hypothesis of uniformity of the distribution is less than 0.05. In addition to this, another more severe distribution-testing criterion has been established. For any j ($1 \leqslant j \leqslant k$) the mean number of random numbers occurring in the ith interval ($1 \leqslant i \leqslant 20$) must not lie beyond the limits

$$500 - 2\sqrt{\frac{500}{j}} \leqslant \sum_{p=1}^{j} \frac{\nu_p(i)}{j} \leqslant 500 + 2\sqrt{\frac{500}{j}}$$

(for any i and j), where $\nu_p(i)$ is the number of objects occurring in the ith interval at the pth test.

The latter criterion is based on a trivial probability relation: with probability 0.95 we have

$$\left| \sum_{p=1}^{j} \frac{\nu_p(i)}{j} - 500 \right| \leqslant 2\frac{\sigma}{\sqrt{j}},$$

where σ^2 is the overall variance of the number of elements in the ith interval. However

$$\sigma^2 = npq = 10{,}000 \cdot \frac{1}{20} \cdot \frac{19}{20} \approx 500 \; (\sigma \sim \sqrt{500}).$$

After k tests, the distribution is reckoned to be uniform if and only if both criteria are satisfied at the same time.

A suggested criterion for testing the distribution in the digits of a random-number generator is the estimate of the distribution of the empirical entropy with respect to all digits.

Let k be the number of binary digits of the random-number generator and let m_i be the number of zeros obtained in n tests in the ith digit. Let us denote by H_i the empirical entropy in the ith digit

$$H_i = -\frac{m_i}{n}\ln\frac{m_i}{n} - \left(1 - \frac{m_i}{n}\right)\ln\left(1 - \frac{m_i}{n}\right).$$

It can be shown that if the probability of occurrence of a zero in all digits is equal to 0.5, then the random quantity

$$\zeta = 2n\left(k\ln 2 - \sum_{i=1}^{k} H_i\right)$$

is asymptotically distributed, for $n \to \infty$, as χ_k^2.

Tests of the distribution of the entropy ζ must be carried out for a sufficiently large n (for example $n = 10{,}000$). The empirical distribution of the values of ζ obtained is compared with the χ_k^2 distribution. If the comparison gives satisfactory results, the distribution of the digits of the random-number generator is reckoned to be sufficiently close to uniform.

c) The execution of a test associated with evaluation of the parameters of a control problem is obligatory. We can use as such a control problem the evaluation of the volume of an n-dimensional hypersphere.

When the results of all tests are satisfactory the random-number generator may be assumed to be working normally and to be ready for operation. If some test gives unsatisfactory results, operation of the generator must be investigated more closely. Several other statistical characteristics have to be calculated. These are mainly data concerning the spectral structure of the random process of the generation of numbers by the random-number generator.

The production by a random-number generator of random numbers uniformly distributed in the interval (0, 1) can be considered as a stationary random process $x(t)$ with discrete time.

At any instant of time the random quantity $x(t)$ has one and the same uniform distribution. Therefore the expectation $\mathsf{M}x(t)$ and the variance $\mathsf{D}x(t)$ are constant for any instant of time and are equal to $^1/_2$ and $^1/_{12}$ respectively.

It is well known that a probability process $x(t)$ is called stationary in the broad sense if

$$\mathsf{M}[x(t)^2] < \infty$$

for all values of t and if

$$B(\tau) = \mathsf{M}\{[x(t+\tau) - \mathsf{M}x(t)][x(t) - \mathsf{M}x(t)]\}$$

does not depend on t. The function $B(\tau)$ is called the covariance function of the process.

In the case of the random process of obtaining numbers from a random-number generator all these conditions apply and, moreover, the covariance function $B(\tau)$ need not depend on τ but has a constant value equal to zero.

The determination and testing of statistical estimates of the random process are conveniently carried out by averaging data obtained as a result of a concrete realization of the random process over a sufficiently long interval of time.

It is well known that the variance of the mean value $\bar{x}(t)$ is equal to $\frac{\mathsf{D}x(t)}{n}$, where n is the number of realizations on which averaging is carried out. From this, confidence intervals for the sampling average $\bar{x}(t)$ can easily be constructed.

We have, with probability $\Phi(k) = \frac{2}{\sqrt{2\pi}} \int_0^k e^{-\frac{t^2}{2}} dt$, for $n = 1000$:

$$\mathsf{P}\left\{\frac{1}{2} - \frac{k}{\sqrt{1000 \cdot 12}} \leqslant \bar{x}(t) \leqslant \frac{1}{2} + \frac{k}{\sqrt{1000 \cdot 12}}\right\};$$

for $p = 0.997$ we have $k = 3$.

Therefore, with probability 0.997 we have

$$0.47 \leqslant \bar{x}(t) \leqslant 0.53.$$

For $k = 2$, $p = 0.95$ and with probability 0.95 we have

$$0.48 \leqslant \bar{x}(t) \leqslant 0.52.$$

A statistical estimate of the covariance function $B(\tau)$ is given by the formula

$$B(\tau) = \frac{1}{n} \sum_{i=1}^{n} [(x(t_i) - \bar{x})(x(t_{i+\tau}) - \bar{x})],$$

where $x(t_i)$ is the value of the realization of the random number at the time of the ith step of operation of the electronic computer, and $\bar{x}$ is the mean value of the random numbers (~ 0.5).

A very useful test to check the quality of a random-number generator is to compute the mutual covariance function $B_{xy}(\tau)$ for the stationary processes of generating zeros and ones in the various digits of the random number.

If the contents of the digits in random numbers generated by means of random-number generators depend on each other, the function $B_{xy}(\tau)$ will either vary with time or else will assume values differing from those calculated theoretically.

The quantity $B_{xy}(\tau)$ has the following estimate

$$B_{xy}(\tau)=\frac{1}{n}\sum_{i=1}^{n}[x(t_i)-\bar{x}(t)][y(t_{i+\tau})-\bar{y}(t)],$$

where $x(t_i)$ and $y(t_{i+\tau})$ are the contents of the two tested digits of the random number generated at the jth step of operation of the electronic computer, $\bar{x}(t)$ and $\bar{y}(t)$ have the value 0.5, and n is the size of the sampled sequence ($n \approx 1000$).

A confirmation of absence of correlation is provided by a statistical estimate of the correlation coefficient between digits. The calculation of the empirical value of the correlation coefficient r is carried out from sampling data of size n according to the formula

$$r=\frac{\frac{1}{n}\left[\sum_{i=1}^{n} x_i, y_i-\frac{\left[\sum_{i=1}^{n}(x_i)\right]\left[\sum_{i=1}^{n}(y_i)\right]}{n}\right]}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2-\bar{y}^2\right)}}.$$

To establish whether correlation coefficients r obtained from observations are significant or whether their departure from zero may be explained as the result of random sampling, we must test the hypothesis $\rho=0$ (ρ is the theoretical correlation coefficient). The empirical correction coefficient r is used as the test statistic. By choosing a confidence coefficient p, we construct a region of the form

$$|r|>t_p\sigma_r,$$

and if the value of the empirical correlation coefficient r obtained lies beyond these limits, then the hypothesis $p=0$ is rejected.

Here t_p is determined from the relation

$$\frac{2}{\sqrt{2\pi}}\int_0^{t_p} e^{-\frac{z^2}{2}}\,dz = p,$$

and

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}.$$

In other words, if

$$|r| > t_p\frac{1-r^2}{\sqrt{n}},$$

we reckon that correlation exists between the digits. The value $p = 0.99$ is taken for the confidence coefficient p. Hence $t_p = 2.58$.

This type of test must form an integral part of the overall program for testing the operation of random-number generators.

Chapter VII

Transformation of Random Numbers

1. PROPERTIES OF QUASI-UNIFORM QUANTITIES

The construction of algorithms for the solutions of problems by the Monte Carlo method is intimately connected with the design of random experiments. In this connection there often arises the need for generating random quantities or processes of various types. These comprise in the first instance independent and dependent random events, Markov chains, random quantities with assigned distribution laws, random vectors, stationary and nonstationary random processes with assigned probability characteristics, etc.

It has been shown in Chapter VI how to generate sequences of uniformly distributed random quantities either by means of special subprograms (in which case pseudo-random numbers are obtained) or by means of random-number generators.

In this chapter we describe algorithms for transforming uniformly distributed quantities, enabling us to obtain random quantities or random vectors with practically any distribution law. Thus the generation of random quantities on electronic computers reduces to obtaining independent quantities with uniform distribution laws and to subsequently transforming them by means of special programs. We could also use special generators for quantities with various distribution laws (for example with the Gaussian or Poisson's distribution law), but this approach is not followed in practice in general-purpose electronic computers.

Thus, we shall assume that we have available a set $\{R_i\}$ $(i = 1, 2, 3, \dots)$ of random numbers R_i having uniform distribution over the interval $(0,1)$. This is called the initial set. On the numbers R_i we need to carry out such operations that we may obtain, as a result, a set $\{S_i\}$ of values of a random quantity with an assigned distribution law. In view of the fact that the number of digits of the numbers processed by electronic computers is finite, we shall not be able to operate, strictly speaking, with random numbers

R_i having uniform distribution over the interval (0,1); what we shall actually have is a set $\{R_i^*\}$ of random numbers R_i^* having a so-called quasi-uniform distribution.

As is well known, a continuous random quantity ξ uniformly distributed over the interval (0,1) has the density function

$$f_\xi(x)=\begin{cases} 1 & \text{for } 0\leqslant x\leqslant 1, \\ 0 & \text{for } x<0 \text{ and } x>1. \end{cases} \tag{7.1}$$

The expectation of the random quantity is equal to

$$\mathsf{M}[\xi]=\frac{1}{2}, \tag{7.2}$$

the variance is equal to

$$\mathsf{D}[\xi]=\frac{1}{12}, \tag{7.3}$$

and the standard deviation is equal to

$$\sigma_\xi=\frac{1}{2\sqrt{3}}. \tag{7.4}$$

The random numbers R_i^* are realizations of a random quantity ζ differing from ξ. Let us study the distribution of ζ. The random numbers R_i^* are so generated (see Chapter VI) that at each digit of a k-digit number there occurs a zero or a one, each with probability equal to 0.5. Therefore possible values of the discrete random quantity ζ will be numbers R_i^*

$$0,\ \frac{1}{2^k},\ \frac{2}{2^k},\ \dots,\ \frac{i}{2^k},\ \dots,\ \frac{2^k-1}{2^k},$$

the probability p_i of each of which is equal to 2^{-k}.

The expectation of the quantity ζ is equal to

$$\mathsf{M}\zeta=\sum_{i=0}^{2^k-1}\frac{i}{2^k}\cdot\frac{1}{2^k}=\frac{1}{2^{2k}}\sum_{i=0}^{2^k-1} i=\frac{2^k(2^k-1)}{2^{2k+1}}=\frac{1}{2}\left(1-\frac{1}{2^k}\right). \tag{7.5}$$

The variance is easily evaluated by proceeding from the expression

$$\zeta=\sum_{l=1}^{k}\varepsilon_l 2^{-l}. \tag{7.6}$$

where ε_l are independent Bernoullian quantities assuming with equal probability the values zero and unity.

Thus the variance of the quantity ζ is equal to

$$D\zeta=\sum_{l=1}^{k}2^{-2l}D\varepsilon_l=\frac{1}{4}\sum_{l=1}^{k}2^{-2l}=\frac{1}{4\cdot 4}\frac{1-\frac{1}{4^k}}{1-\frac{1}{4}}=\frac{1}{12}\left(1-\frac{1}{4^k}\right). \tag{7.7}$$

The departure of the expectation (7.5) from the value 1/2 can prove significant if the number k of digits is small. Therefore it is often convenient to consider a quantity η of the form

$$\eta=\frac{2^k}{2^k-1}\zeta. \tag{7.8}$$

This quantity assumes the values

$$0,\ \frac{1}{2^k-1},\ \frac{2}{2^k-1},\ \dots,\ 1.$$

The expectation of the random quantity η is equal to

$$M[\eta]=\sum_{i=0}^{2^k-1}\frac{i}{2^k-1}2^{-k}=\frac{2^{-k}}{2^k-1}\sum_{i=0}^{2^k-1}i.$$

However

$$\sum_{i=0}^{2^k-1}i=\frac{2^k(2^k-1)}{2},$$

and therefore

$$M[\eta]=\frac{1}{2}. \tag{7.9}$$

The variance is equal to

$$D[\eta]=\sum_{i=0}^{2^k-1}\left(\frac{i}{2^k-1}-\frac{1}{2}\right)^2 2^{-k}.$$

By taking into account that

$$\sum_{i=0}^{2^k-1}i^2=\frac{(2^k-1)\cdot 2^k(2^{k+1}-1)}{6},$$

we shall obtain

$$\mathrm{D}[\eta]=\frac{1}{12}\,\frac{2^k+1}{2^k-1}. \tag{7.10}$$

The standard deviation will be equal to

$$\sigma_\eta=\frac{1}{2\sqrt{3}}\sqrt{\frac{2^k+1}{2^k-1}}. \tag{7.11}$$

When $k \to \infty$ we have $\sigma_\eta \to \frac{1}{2\sqrt{3}}$. The order of approximation can be seen from Table 14.

Table 14

k	$\mathrm{D}[\eta]$	σ_η	k	$\mathrm{D}[\eta]$	σ_η	k	$\mathrm{D}[\eta]$	σ_η	k	$\mathrm{D}[\eta]$	σ_η
2	0.1389	0.3727	5	0.0887	0.2979	8	0.0840	0.2898	12	0.0834	0.2888
3	0.1071	0.3274	6	0.0860	0.2933	9	0.0837	0.2893	15	0.08334	0.28870
4	0.0945	0.3073	7	0.0846	0.2910	10	0.0835	0.2889	20	0.08333	0.28868

We observe that $\mathrm{D}[\xi]=\frac{1}{12}\approx 0.08333$ and $\sigma_\xi=\frac{1}{2\sqrt{3}}\approx 0.28868$.

If the number of digits of the numbers processed in the computer is large, the difference between $\{R_i^*\}$ and uniformly distributed random numbers may be ignored in the solution of the majority of computing problems. Problems exist, however, for the solution of which this difference must be taken into account.

Below, unless otherwise stated, we shall assume the initial set of numbers $\{R_i\}$ to be uniformly distributed over the interval (0,1).

2. SIMULATION OF INDEPENDENT RANDOM EVENTS

Suppose we have to simulate a random event A which occurs with given probability p. Let ξ be a random quantity having a uniform distribution over the interval (0,1). The values assumed by this quantity are denoted, as above, by R_i.

We shall identify the event A with the event "a selected value R_i of the random quantity ξ satisfies the inequality

$$R_i \leqslant p.\text{"} \tag{7.12}$$

It can easily be seen that the probability of the event A is equal to

$$P(A)=\int_0^p dx=p. \qquad (7.13)$$

Then the opposite event $\bar{A}$ is the event

$$R_i > p$$

and its probability is equal to $P(\bar{A})=1-p$.

A procedure for simulating trials of the type considered consists in selecting values R_i and comparing them with the quantity p. If in a given comparison the condition (7.12) is satisfied, then the outcome of the trial is the event A. If the condition (7.13) is not satisfied, the outcome of the test is the event $\bar{A}$.

The above considerations can be extended to a group of events.

Let $A_1, A_2, \ldots, A_s$ be an exhaustive set of events occurring with probabilities $p_1, p_2, \ldots, p_s$. As is well known, we have in this case

$$p_1+p_2+\ldots+p_s=1.$$

We shall identify the event A_m with the event "the selected value R_i of the random quantity ξ satisfies the inequality

$$l_{m-1}<R_i\leqslant l_m," \qquad (7.14)$$

where

$$l_r=\sum_{i=1}^{r} p_i. \qquad (7.15)$$

Similarly to (7.13) we can write

$$P(A_m)=\int_{l_{m-1}}^{l_m} dx=p_m.$$

The trial simulation procedure consists in this case in sequentially comparing random numbers R_i with the quantities l_r. The outcome of a trial proves to be the event A_m if the condition (7.14) is satisfied.

This procedure is sometimes called the determination by lot of the outcome of a trial in which the possible results have probabilities $p_1, p_2, \ldots, p_s$. It is often necessary to examine trials of this kind, each of which is, however, a composite event depending on two or more simple events.

For example, let independent events A and B have probabilities p_A and p_B respectively. Possible outcomes of simultaneous trials will be, in this case, the events

$$AB, \quad \bar{A}B, \quad A\bar{B}, \quad \bar{A}\bar{B} \tag{7.16}$$

with probabilities

$$p_A p_B, \quad (1-p_A)p_B, \quad p_A(1-p_B), \quad (1-p_A)(1-p_B). \tag{7.17}$$

It is evident that to simulate simultaneous trials, two versions of the procedure can be used. The first of them consists in sequentially testing a condition similar to (7.12) with respect to the events A and B. However, we can also proceed differently. A second version may be constructed by analogy with (7.14) as the determination by lot of one of the outcomes (7.16) in accordance with the probabilities (7.17).

The first of the versions considered requires the use of two numbers R_i and two comparisons to test conditions (7.12). In the second version we can restrict ourselves to a single number R_i, but a greater number of comparisons may in general be required. In the practical solution of problems the choice of one version of the procedure or the other is determined by consideration of the ease with which the algorithm may be constructed as well as economy in the number of machine operations required and in the number of locations of the working memory. On the whole the first version proves more economical than the second one.

Let us consider the case when the events A and B are not independent. As before, let the probabilities of the events A and B be denoted by p_A and p_B. In addition we assume that we are given the probability $p(B/A)$ of the event B on the assumption that the event A has occurred.

The first version of the above-mentioned procedure will have, in this case, the following setup.

From the set $\{R_i\}$ there is extracted a number R_n and the validity of the inequality

$$R_n \leqslant p_A \tag{7.18}$$

is tested.

If the inequality (7.18) is found to be valid, this means that the event A has occurred. Therefore for the test connected with the event B we use the probability $p(B/A)$. From the set $\{R_i\}$ we take the next number R_{n+1} and test the condition

$$R_{n+1} \leqslant p(B/A). \tag{7.19}$$

Depending on whether the inequality (7.19) is valid or not, the outcome of the test is either AB or $A\bar{B}$.

If the inequality (7.18) is not found to be valid, this means that the event $\bar{A}$ has occurred. Therefore, for the test connected with the event B we must use the probability $p(B/\bar{A})$. This probability can be determined from the theorem of total probability

$$p(B) = p(A)\,p(B/A) + p(\bar{A})\,p(B/\bar{A}),$$

from which it follows that

$$p(B/\bar{A}) = \frac{p_B - p_A p(B/A)}{1 - p_A}. \tag{7.20}$$

We select from the set $\{R_i\}$ the number R_{n+1} and test the validity of the inequality

$$R_{n+1} \leqslant p(B/\bar{A}).$$

Depending on whether this inequality is valid or not, we obtain for the test the results $\bar{A}B$ or $\bar{A}\bar{B}$.

We can also use the second version of simulating procedure. To do this it suffices to note that the events

$$AB,\quad A\bar{B},\quad \bar{A}B,\quad \bar{A}\bar{B}$$

form an exhaustive set and have probabilities

$$\begin{gathered} p_A \cdot p(B/A), \quad p_A[1 - p(B/A)], \\ (1 - p_A)\,p(B/\bar{A}), \quad (1 - p_A)[1 - p(B/\bar{A})], \end{gathered}$$

where $p(B/\bar{A})$ is determined by the relation (7.20).

More complex algorithms can also be constructed in a similar manner. We shall restrict ourselves to showing the principle of simulation of simple Markov chains.

A simple homogeneous Markov chain is determined by a transition matrix

$$\Pi = \begin{vmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{vmatrix}.$$

Possible results of trials are the events $A_1, A_2, \ldots, A_k$. The probability p_{ij} is the conditional probability of occurrence of the event A_j in a trial, given that the result of the previous trial was the event A_i.

The simulation of such a Markov chain consists in sequentially selecting events A_j by lot according to probabilities p_{ij}. It proceeds as follows.

We choose first an initial state assigned by the initial probabilities $p_{01}, p_{02}, \ldots, p_{0k}$.

To do this we select from the set $\{R_i\}$ a number R_n and compare it with the quantities l_r as in (7.14), where the quantities $p_{01}, p_{02}, \ldots, p_{0k}$ are used as the p_i's in (7.15).

In this way we determine an integer m_0 which satisfies the inequality $l_{m_0} < R_n \leqslant l_{m_0+1}$.

Thus the initial event of the realization of the chain will be the event A_{m_0}. Then we select the next random number R_{n+1}, which is also compared with quantities l'_r. Here, however, we use, as the probabilities p_i for determining l'_r, the elements $p_{m_0 1}, p_{m_0 2}, \ldots, p_{m_0 k}$ of the transition matrix. By means of the comparison, the number m_1 for which the condition tested is verified is established, and the next event of the realization of the chain will be the event A_{m_1}. We continue in a similar manner. It is evident that each number m_i determines not only the next event A_{m_i} of the realization but also the distribution of the probabilities $p_{m_i 1}, p_{m_i 2}, \ldots, p_{m_i k}$ for selecting the next integer m_{i+1}.

We observe that for ergodic Markov chains the effect of the initial probabilities decreases rapidly with the increase of the ordinal number of the test. Therefore we can take, for the quantities $p_{01}, p_{02}, \ldots, p_{0k}$, arbitrary values, for example

$$p_{01} = p_{02} = \ldots = p_{0k} = \frac{1}{k}.$$

The key idea of the procedure described is retained also for more complicated Markov chains, for example, for nonhomogeneous chains.

3. TYPICAL FEATURES OF THE SIMULATION OF EVENTS BY MEANS OF RANDOM NUMBERS WITH FEW DIGITS

The rules considered above are strictly valid only in the case when random numbers R_i having uniform distribution over the interval (0,1) are used to simulate trials.

However, when electronic digital computers are used, only random numbers R_i^* with quasi-uniform distribution are available. This case has typical features which it is desirable to discuss here, if only briefly.

Let us consider first the determination of results of trials by verifying the validity of condition (7.12).

Suppose we have available k-digit numbers with possible values

$$R_i^* = \frac{i}{2^k - 1} \qquad (i = 0,\ 1,\ 2,\ \ldots,\ 2^k - 1). \tag{7.21}$$

We now replace R_i in the inequality (7.12) by the number R_i^*. Thus, the simulated event A^* is identified with the event

$$R_i^* \leqslant p. \tag{7.22}$$

The probability $P(A^*)$ can be found as the ratio of the number n of numbers of the form (7.21) smaller than or equal to p, to the number N of all numbers of the form (7.21). As is well known, $N = 2^k$. Thus

$$P(A^*) = \frac{n}{2^k}. \tag{7.23}$$

We can draw from the relation (7.23) the following conclusion: if the probability p of the event A is comprised within the limits

$$\frac{n}{2^k} \leqslant p < \frac{n+1}{2^k}, \tag{7.24}$$

then $P(A^*) = \frac{n}{2^k}$.

It follows from this that the use of the numbers R_i^* instead of R_i leads to an error in the value of the probability of the event, equal to $\frac{n}{2^k} - p = \Delta p$. Clearly the maximum value of the error Δp does not exceed $\frac{1}{2^k - 1}$.

A similar situation arises also in the more complex case in which results of trials are determined by verifying the validity of conditions (7.14).

4. METHODS OF OBTAINING RANDOM NUMBERS WITH ASSIGNED DISTRIBUTION LAW

There exists a basic relation that connects random numbers S_i having an assigned distribution law and numbers R_i having uniform distribution over the interval (0,1).

The possibility of obtaining such a relation derives from the following theorem (see, for example, [33]): if the random quantity ξ has a probability density $f(x)$, then the distribution of the random quantity

$$\eta = \int_{-\infty}^{\xi} f(x)\,dx \tag{7.25}$$

is uniform over the interval (0,1).

On the basis of this theorem we can establish the following rule. In order to obtain a number belonging to the set of random numbers $\{S_i\}$ having the probability density function $f(x)$, we must solve with respect to S_i the equation

$$\int_{-\infty}^{S_i} f(x)\,dx = R_i. \tag{7.26}$$

To prove this rule, we consider a random quantity η having uniform distribution over the interval (0,1), and a random quantity ξ connected with η by the relation (7.25).

We assume that $f(x)$ nowhere reduces identically to zero. Then, accordingly to (7.25), η is a monotonically increasing function of ξ, and therefore ξ can be expressed as a single-valued function of η

$$\xi = \varphi(\eta). \tag{7.27}$$

It can easily be seen that the inverse function

$$\eta = \varphi^{-1}(\xi)$$

is expressed in this case by the relation (7.25). By bearing this in mind we shall find the probability density of the random quantity ξ. The distribution function $F_\xi(x)$ is equal to the probability that $\xi < x$:

$$F_\xi(x) = P(\xi < x). \tag{7.28}$$

We substitute for ξ in (7.28) its value from (7.27)

$$F_\xi(x) = P[\varphi(\eta) < x]. \tag{7.29}$$

Since the function $\xi = \varphi(\eta)$ is monotonically increasing, the inequality

$$\varphi(\eta) < x$$

is equivalent to the inequality

$$\eta < \varphi^{-1}(x).$$

Therefore

$$F_\xi(x) = P[\eta < \varphi^{-1}(x)]. \tag{7.30}$$

The probability occurring in (7.30) can be evaluated since the density function $f_\eta(y)$ of the random quantity η is known: we have in fact assumed that η is uniformly distributed over the interval (0,1). Therefore

$$F_\xi(x) = \int_0^{\varphi^{-1}(x)} f_\eta(y)\,dy$$

or

$$F_\xi(x) = \int_0^{\varphi^{-1}(x)} dy = \varphi^{-1}(x). \tag{7.31}$$

By substituting for $\varphi^{-1}(x)$ in (7.31) its expression from (7.25) we obtain

$$F_\xi(x) = \int_{-\infty}^{x} f(x)\,dx.$$

The latter relation shows that the random quantity has the density function $f(x)$.

In a number of cases, the relation (7.26) can be directly used for practical purposes. Let us consider some examples.

Suppose it is required to obtain random numbers with an exponential distribution law

$$f(x) = \lambda e^{-\lambda x} \qquad (x > 0). \tag{7.32}$$

By virtue of relation (7.26) we obtain

$$\lambda \int_0^{x_i} e^{-\lambda x}\,dx = R_i$$

or, after evaluating the integral,

$$1 - e^{-\lambda x_i} = R_i.$$

By solving this equation with respect to x_i, we obtain

$$x_i = -\frac{1}{\lambda} \ln (1 - R_i). \tag{7.33}$$

If we have available random numbers R_i uniformly distributed over the interval (0,1), then, by using formula (7.33), we can construct a sequence of random numbers x_i having the exponential distribution (7.32).

Suppose it is required to obtain random numbers x_i with distribution law

$$f(x)=\lambda\left(1-\frac{\lambda}{2}x\right) \quad \left(0\leqslant x\leqslant\frac{2}{\lambda}\right), \tag{7.34}$$

which finds application in the solution of certain problems of queueing theory.

By using relation (7.26) we can obtain

$$\lambda\left(x_i-\frac{\lambda}{4}x_i^2\right)=R_i.$$

Hence

$$x_i=\frac{2}{\lambda}(1-\sqrt{1-R_i}).$$

As our next example, let us consider the probability density function

$$f(x)=\frac{c}{(1+bx)^2} \quad \left(0\leqslant x\leqslant\frac{1}{c-b}\right).$$

Relation (7.26) has in this case the form

$$\frac{cx_i}{1+bx_i}=R_i.$$

Therefore

$$x_i=\frac{R_i}{c-bR_i}. \tag{7.35}$$

It may be shown in a similar manner that the quantities

$$S_i=\sigma\sqrt{-\ln R_i}\,\sqrt{2}$$

are distributed according to the Rayleigh law with parameter σ:

$$F(x)=\begin{cases} 1-e^{-\frac{x^2}{2\sigma^2}}, & x\geqslant 0\\ 0, & x<0.\end{cases}$$

Random numbers distributed according to this law play an important role in radio-engineering problems.

In certain problems we have to operate with random quantities η distributed according to Pareto's law

$$F(x)=\begin{cases} 1-\frac{1}{x^{\alpha}}, & x \geqslant 1, \\ 0, & x < 1. \end{cases}$$

Such quantities are obtained from uniformly distributed pseudo-random numbers ξ by means of the formula

$$\eta=\xi^{-\frac{1}{\alpha}},$$

which is of particularly simple form if $\alpha=\frac{1}{k}$, where k is an integer.

We have considered above examples of transformation of random numbers into possible values of continuous random quantities.

We consider now methods of simulating discrete random quantities. Suppose a discrete random quantity ξ assumes values x_i with probabilities $P(x=x_i)=p_i$. We call the occurrence of a possible value x_i the event A_i. If the number of x_i's is finite, we can directly use the methods of the previous section. These methods are often applicable, however, also for an infinite set of values x_i.

Suppose, for example, that we need to obtain random numbers having Poisson's distribution

$$P(n)=\frac{a^n}{n!}e^{-a} \qquad (n=0, 1, 2, \ldots). \tag{7.36}$$

To do this we select random numbers R_i and test the validity of inequalities of the type (7.14)

$$l_{n-1} < R_i \leqslant l_n, \tag{7.37}$$

where

$$l_r=e^{-a}\sum_{n=0}^{r}\frac{a^n}{n!} \qquad (r=0, 1, 2, \ldots); \quad l_{-1}=0. \tag{7.38}$$

If the inequalities (7.37) are found to be satisfied, the next random number η is assumed equal to n. Clearly the quantity η has the Poisson distribution.

Rules for the transformation of random numbers, based on relation (7.26), enable us to obtain a sequence of random numbers x_i that satisfy exactly an assigned distribution law if and only if the numbers R_i are uniformly distributed over the interval (0,1).

In practice, instead of the numbers R_i, we usually employ random numbers R_i^*, having quasi-uniform distribution. Owing to this the assigned distribution law is only approximately realized.

If the number of digits of the numbers R_i^* is equal to k, the transformation defined by the relation (7.26) operates on 2^k distinct numbers R_i^*. Therefore the number of distinct values x_i cannot exceed 2^k. If the numbers x_i must be represented with k digits, then to represent all 2^k numbers x_i we may require in general a number of digits $k^* > k$. This is explained both by the nonuniform disposition of the numbers x_i on the numerical axis and by the fact that the numbers x_i may exceed in absolute value the limits represented by k-digit numbers.

Let us proceed to consider methods of transformation of random numbers not connected with the solution of equation (7.26).

One of them consists in selecting, from a uniform set $\{R_i\}$, random numbers x_j satisfying a certain condition in such a manner that x_j will be subject to the assigned distribution law.

Suppose it is required to obtain random numbers x_i having the probability density function $f_\xi(x)$. If the region of possible values of the random quantity ξ is not restricted on both sides, we have to pass to the corresponding truncated distribution. We assume that the region of possible values for the truncated distribution is the interval (a, b).

By the change of variables

$$\eta = \frac{\xi - a}{b - a} \tag{7.39}$$

we pass to the random quantity η with probability density function

$$f_\eta(y) = (b - a) f_\xi [a + (b - a) y]. \tag{7.40}$$

The random quantity η has the interval (0,1) as its region of possible values. Let the maximum value of $f_\eta(y)$ in this interval be equal to f_m. By a change of scale on the axis $[0, f_\eta(y)]$ we reduce the interval $(0, f_m)$ to a length equal to unity. Then

$$f_\eta^*(y) = \frac{b - a}{f_m} f_\xi [a + (b - a) y]. \tag{7.41}$$

The resulting probability density function $f^*_\eta(y)$ may be inscribed in a unit square as is shown in Fig. 42.

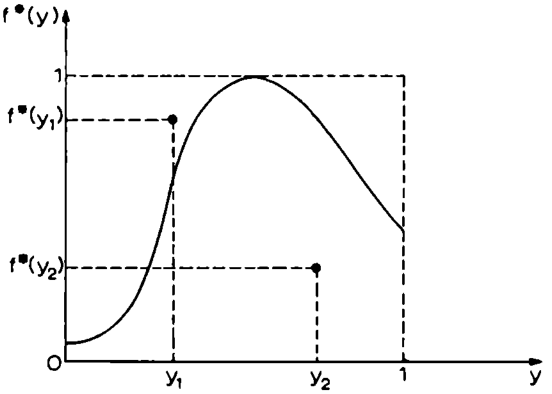


Fig. 42

Let us choose a pair of numbers R_{2l-1}, R_{2l} from a sequence of uniformly distributed quantities. This pair determines a random point in the unit square (Fig. 42).

We shall take as values y_l only those numbers $\tilde{R}_{2l}$ of the set $\{R_{2l}\}$ that satisfy the condition

$$R_{2l} \leqslant f^*_\eta(R_{2l-1}) \tag{7.42}$$

or

$$R_{2l} \leqslant \frac{b-a}{f_m} f_\xi\,[a+(b-a)\,R_{2l-1}]. \tag{7.43}$$

There emerges thus the following procedure for generating a random quantity having probability density $f_\xi(x)$.

From a set of random numbers R_l uniformly distributed over the interval (0,1) we select pairs of numbers R_{2l-1}, R_{2l}. We test for the numbers R_{2l-1}, R_{2l} the validity of the inequality (7.43). If (7.43) is found to be satisfied, the next number x_j is determined from the relation

$$x_j = a+(b-a)\,R_{2l}. \tag{7.44}$$

If the inequality (7.43) is not satisfied, the pair of numbers R_{2l-1}, R_{2l} is rejected and the following pair is selected.

It can be seen that the random numbers x_i thus obtained have the distribution density $f_\xi(x)$.

The methods given above for transforming random numbers require for their realization a considerable number of operations. This is especially true in those cases in which either Eq. (7.26) cannot be solved exactly or else the calculation of the right-hand side of the inequality (7.43) proves to be very laborious. In particular, none of these methods is applicable in practice for obtaining a sequence of random numbers with the Gaussian distribution.

Let us consider methods of transforming random numbers which are based on approximate simulation of the conditions for which the limit theorems of probability theory are valid.

Suppose it is required to obtain a set of random numbers $\{S_i\}$ having a Gaussian distribution with expectation a and standard deviation σ,

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (7.45)$$

By virtue of the central limit theorem of probability theory and under some very general conditions, the sums of a large number of random components are asymptotically distributed according to the Gaussian law. Therefore we can use the addition of numbers of an initial set for approximately simulating normally distributed random numbers.

Since in this case the components have uniform distribution over the interval (0,1), we find from the limit theorem for identically distributed random quantities that, if the independent random quantities $\xi_1, \xi_2, \ldots, \xi_k$ have all one and the same probability distribution and if each ξ_i has expectation a^* and standard deviation σ^*, then the sum

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n$$

is asymptotically normal with expectation

$$a = a^* n$$

and standard deviation

$$\sigma = \sigma^* \sqrt{n}.$$

It can be shown that the sum ξ has a distribution close to normal even for comparatively small values of n. In the solution of applied problems n is usually taken equal to 4—8.

As is well known, for a random quantity having uniform distribution over the interval (0,1) the expectation is equal to 1/2 and the standard deviation is $\frac{1}{2\sqrt{3}}$. Therefore the sum of n such quantities will have expectation

$$a=\frac{n}{2}\tag{7.46}$$

and standard deviation

$$\sigma=\frac{1}{2}\sqrt{\frac{n}{3}}.\tag{7.47}$$

Let us consider now what happens when, to obtain normally distributed random numbers by summation, we use quasi-uniform numbers R_i^*.

It has been shown in Section 1 that random numbers having quasi-uniform distribution over the interval (0,1) are characterized by expectation 1/2 and standard deviation

$$\sigma^*=\frac{1}{2\sqrt{3}}\sqrt{\frac{2^k+1}{2^k-1}}.$$

Therefore the expectation of the sum ξ is given by expression (7.46) and the standard deviation is

$$\sigma=\frac{\sqrt{n}}{2\sqrt{3}}\sqrt{\frac{2^k+1}{2^k-1}}.\tag{7.48}$$

The need for relation (7.48) arises only when random numbers R_i^* with few digits are used in the summation process.

Of course, an increase of the number n of components in the sum ξ leads to the distribution law of ξ coinciding more closely with a normal distribution. However, this also leads to an increase of the number of operations needed for transforming random numbers.

As has been shown in [7], special transformations may be used to improve the asymptotic normality of the sum ξ.

For example, if η is the normalized sum

$$\eta=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\tag{7.49}$$

of random quantities ξ_i having uniform distribution over the interval $(-h, +h)$, then the quantity

$$\xi=\eta-\frac{1}{20n}(3\eta-\eta^3)\tag{7.50}$$

will have a distribution sufficiently close to normal for values of n substantially smaller than in the case of simple summation. The distribution law of the quantity ξ will be close to normal for n as small as 5 (see [7]). If another transformation is used, namely

$$\xi^* = \eta - \frac{41}{13400n^2}(\eta^5 - 10\eta^3 + 15\eta), \tag{7.51}$$

then it suffices for practical purposes to have $n = 2$.

To use this type of transformation a larger number of operations is required than in simple summation. Therefore, only by estimating the expenditure of machine time for solving a given class of problems can we assess the expediency of choosing a particular value of n for the number of components and using one transformation rather than another.

Random normal numbers can be used to obtain functions of normally distributed quantities. As an example, we shall give the method of generating numbers distributed according to Rice's law

$$p(x) = \frac{x}{\sigma} e^{-\frac{x^2+s^2}{2\sigma^2}} I_0(sx),$$

where $I_0(x)$ is the modified Bessel function of the first kind and zero order.

The general method of obtaining numbers on the basis of inverting the distribution function is obviously inapplicable here. An indirect approach is based on the fact that Rice's law describes the probability distribution of the absolute value (the modulus) of a random vector on a plane, with independent normally distributed components having variance σ^2 and nonzero expectation. To construct a concrete example, we can assume that the expectation of the first component is equal to S and that of the second component is equal to zero. By denoting these normal quantities by ξ_1 and ξ_2, we express the required Rice quantity η in terms of them:

$$\eta = \sqrt{\xi_1^2 + \xi_2^2}.$$

Another method of electronic computer generation of quantities with Rice distribution is based on their physical meaning as the absolute value of the vector sum of two vectors: one constant of length s and one random having length ζ distributed according to the Rayleigh law.* By considering the triangle formed by these vectors and by denoting by φ the angle between the component vectors, we have by the cosine theorem

---

* The first vector simulates the voltage of a constant signal and the second vector a noise voltage.

$$\eta=\sqrt{s^2+\zeta^2-2s\zeta\cos\varphi}.$$

A similar idea can be applied for obtaining random numbers having a Poisson distribution

$$P(k)=\frac{a^k}{k!}e^{-a}$$

with expectation a.

Here we use Poisson's limit theorem: if p_n is the probability of occurrence of an event A in a single trial, the probability of occurrence of k events in n independent trials for $n\to\infty$ and $p_n\to 0$ is asymptotically equal to $P(k)$.

We choose a value of n large enough to ensure that

$$p_n=\frac{a}{n} \tag{7.52}$$

is smaller than unity (it is usually required in practice that $p_n=0.1-0.2$). If we simulate a series of n independent trials in each of which the event A occurs with probability p_n, then the random number x_i having Poisson distribution must be chosen to be the number of cases of the actual occurrence of the event A.

In conclusion, we shall consider a fairly general approximate method of transforming random numbers, based on piecewise approximation to the probability density. This method admits of a very simple realization on computers and in many cases ensures the required transformation accuracy. It has, therefore, wide acceptance in practice.

Suppose we need to obtain random numbers with probability density $f_\xi(x)$ and a bounded region of possible values (c, d). If the range of possible values is not bounded, we pass to the corresponding truncated distribution. Let us split (c, d) into n intervals, and denote by a_k the left-hand boundary of the kth interval.

We can adopt the following procedure for transforming random numbers:

1) the random choice of an interval among the n possible intervals (the determination of the quantity a_k),

2) the random choice of the number y_{ki} in the interval (a_k, a_{k+1}),

3) the generation of a random number x_i from the relation

$$x_i=a_k+y_{ki}. \tag{7.53}$$

From the point of view of realization on computers it is most convenient to choose the probabilities of occurrence in all intervals (a_k, a_{k+1}) to be the same.

In this case we choose a number of intervals n large enough to ensure the required accuracy in the transformation of the random numbers. This number is conveniently chosen in such a manner that $n = 2^m$, where m is an integer smaller than or equal to the number of binary digits of the random numbers of the initial quasi-uniform set.

In the simplest and most common case, the density function $f_\xi(x)$ is approximated within each interval by a constant value f_k. This means that the random quantity η_k has a uniform distribution in the interval (a_k, a_{k+1}).

The quantities f_k and a_k are evaluated in advance and are placed in the memory of the computer before the beginning of the computations connected with the transformation of random numbers. To calculate the values of a_k and f_k we can proceed in the following manner. If the number of intervals (Fig. 43) into which the range (c, d) is divided is equal to n, the probability of occurrence of the random quantity ξ in the interval (a_k, a_{k+1}) must be equal to $1/n$. Therefore

$$\int_{a_k}^{a_{k+1}} f_\xi(x)\, dx = \frac{1}{n}. \tag{7.54}$$

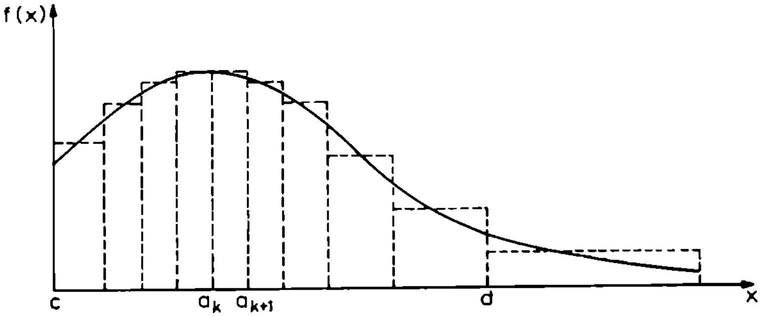


Fig. 43.

Since a_1 is known and is equal to c the relation (7.54) enables us to evaluate one after another all the required values a_k. Now, for each interval we can easily evaluate the values f_k from the relation

$$f_k \cdot (a_{k+1} - a_k) = \frac{1}{n}. \tag{7.55}$$

The fact that the evaluation of the quantities a_k sometimes involves lengthy computations is not very important, since these calculations, being carried out in advance, do not affect the number of operations in the random-number transformation program.

The random-number transformation procedure reduces to the following. From a set of quasi-uniform random numbers we extract a number R^*_{2j-1} and we use the first $m = \log_2 n$ digits of this number as the address for sampling from a table of values of a_k and a_{k+1}. Then we extract the next random number R^*_{2j} and determine x_i from the relation

$$x_i = a_k + (a_{k+1} - a_k) R^*_{2j}. \tag{7.56}$$

This method for the approximate transformation of random numbers is very compact from the point of view of the number of operations required. It proves particularly convenient when we can choose a comparatively small number n (for example 16, 32 or 64). We observe that the number of operations needed for the transformation of random numbers by means of this method does not depend on n, i.e., it does not depend on the accuracy of the approximation to the distribution law. The accuracy of the approximation determines only the size of the table containing the quantities a_k.

A disadvantage of the method is the fact that the accuracy of the approximation of the function $f_\xi(x)$ is not everywhere the same over the range (c, d). The accuracy depends on the quantity f_k and decreases for small values of f_k. Therefore we must choose the number of intervals n with a view to ensuring the desired accuracy over the interval with the smallest value of f_k.

The relation (7.53) enables us to obtain, on a computer, the value of a Rice-distributed quantity by means of two uniform quantities. One of them is needed for generating a Rayleigh distribution for ζ, and the other, to be later multiplied by 2π, is the phase displacement φ of noise with respect to signal. In the programming method of generating uniformly distributed quantities, the two methods described for generating Rice-distributed quantities require approximately the same expenditure of machine time.

5. SIMULATION OF RANDOM VECTORS AND RANDOM FUNCTIONS

In solving problems by the Monte Carlo method, the need often arises for generating samples of random vectors and in particular discrete realizations of stationary and nonstationary random processes. We shall consider in this connection certain methods of random-number transformation which are convenient from the point of view of machine computation.

Let a two-dimensional random vector (ξ, η) be assigned by means of the joint density function $f(x, y)$. Having available a set of random numbers $\{R_i\}$ which have uniform distribution over the interval (0,1), we need to obtain a sequence of realizations of the random vector (ξ, η).

We find the partial density function of the random quantity η

$$f_\eta(y) = \int_{-\infty}^{+\infty} f(x, y)\,dx. \tag{7.57}$$

From the set of random numbers $\{R_i\}$ we choose a number R_{2j-1} and by one of the methods considered in the previous section we determine the number y_j corresponding to its having distribution density $f_\eta(y)$.

Let us now consider the conditional distribution of the random quantity ξ

$$f_\xi(x/y_j) = \frac{f(x, y_j)}{f_\eta(y_j)}. \tag{7.58}$$

From the set of random numbers $\{R_i\}$ we choose the number R_{2j} and determine a number x_j, corresponding to it, having conditional distribution density $f_\xi(x/y_j)$.

It can be seen that the sequence thus obtained of pairs of numbers (x_j, y_j) has the joint density function $f(x, y)$.

Similar relations are obtained in the case of a three-dimensional vector.

Let the joint density function $f(x, y, z)$ of the components ξ, η and ζ be assigned. If random numbers x_i, y_i, z_i have distributions

$$\left.\begin{aligned} f_\zeta(z) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y, z)\,dx\,dy, \\ f_\eta(y/z_i) &= \int_{-\infty}^{+\infty} \frac{f(x, y, z_i)}{f_\zeta(z_i)}\,dx, \\ f_\xi(x/y_i, z_i) &= \frac{f(x, y_i, z_i)}{f_\zeta(z_i)\cdot f_\eta(y_i/z_i)}, \end{aligned}\right\} \tag{7.59}$$

then the components of the random vector (ξ, η, ζ) will have the joint probability density $f(x, y, z)$. The procedure for generating x_i, y_i, z_i is similar to the two-dimensional case.

These methods of generating random vectors are, however, very laborious. If the integrals of the density functions (7.57) and (7.58) or (7.59) are expressible in closed form, relation (7.26) may be used for generating realizations. Otherwise recourse has to be had to approximate methods. The use of approximate methods is made difficult by the fact that in the simulating process we must take into account the dependence of the conditional probability densities on the values of the components actually obtained.

6. THE SIMULATION OF CERTAIN MULTIDIMENSIONAL QUANTITIES

Let us consider a method of simulating n-dimensional random vectors whose components are normally distributed. We denote by $\eta_1, \eta_2, \ldots, \eta_n$ the components of the random vector. The joint probability density of these quantities is equal to

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n D}} e^{-\frac{1}{2}(x\,\mathfrak{M}^{-1}x)},$$

where $\mathfrak{M}$ is the covariance matrix, $X = (x_1, x_2, \ldots, x_n)$ and $D = \det \mathfrak{M}$. The key idea of the method of simulation to be described is that, having generated n independent normally-distributed quantities $\xi_1, \xi_2, \ldots, \xi_n$ with zero expectations and unit variances, we subject them to such a linear transformation A that the resulting quantities $\eta_1, \eta_2, \ldots, \eta_n$ will have the preassigned covariance matrix $\mathfrak{M} = \|m_{ij}\|$. We observe that the property of normality of a distribution is retained under linear transformation. We shall show now how to choose the linear transformation A in such a manner that it converts the quantities $\xi_1, \xi_2, \ldots, \xi_n$ into $\eta_1, \eta_2, \ldots, \eta_n$ with the assigned covariance matrix.

We assume the transformation matrix to be triangular:

$$\begin{aligned}
\eta_1 &= a_{11}\xi_1, \\
\eta_2 &= a_{21}\xi_1 + a_{22}\xi_2, \\
&\cdots\cdots\cdots\cdots\cdots \\
\eta_n &= a_{n1}\xi_1 + a_{n2}\xi_2 + \ldots + a_{nn}\xi_n.
\end{aligned}$$

Its elements a_{ij} will be found from the condition

$$\mathsf{M}\eta_k\eta_l = m_{kl}.$$

According to our assumptions we have

$$\mathsf{M}\xi_k\xi_l = \delta_{kl} = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

and therefore

$$\mathsf{M}\eta_1^2 = a_{11}^2\mathsf{M}\xi_1^2 = a_{11}^2 = m_{11},$$

i.e.,

$$a_{11} = \sqrt{m_{11}}.$$

Let us determine the elements of the second row of the matrix A:

$$\mathsf{M}\eta_1\eta_2 = a_{11}a_{21}\mathsf{M}\xi_1^2 + a_{11}a_{22}\mathsf{M}\xi_1\xi_2 = a_{11}a_{21} = m_{21},$$

i.e.,

$$a_{21} = \frac{m_{12}}{\sqrt{m_{11}}};$$

and then

$$\mathsf{M}\eta_2^2 = a_{21}^2\mathsf{M}\xi_1^2 + 2a_{21}a_{22}\mathsf{M}\xi_1\xi_2 + a_{22}^2\mathsf{M}\xi_2^2 = a_{21}^2 + a_{22}^2 = m_{22},$$

i.e.,

$$a_{22} = \sqrt{\frac{m_{22}m_{11} - m_{12}^2}{m_{11}}}.$$

By proceeding in a similar manner we can determine one after another all elements of the matrix A: after we find its first k rows the $(k+1)$th row is evaluated by means of $k+1$ equations obtained from the conditions

$$\mathsf{M}\eta_l\eta_{k+1} = m_{l,k+1}, \qquad l = 1, \ldots, k.$$

The choice of a transformation with a triangular matrix is advantageous in that, in the first place, such a matrix occupies only $\frac{n(n+1)}{2}$ locations in the memory of an electronic computer (and not n^2 as would be the case if it were not triangular) and, in the second place, the number of arithmetical operations required to transform the numbers $\eta_1, \ldots, \eta_n$ is reduced.

By denoting by $\bar{\xi}=(\xi_1, \ldots, \xi_n)$ a vector the components of which are independent normal pseudo-random numbers, and by $\bar{\eta}=(\eta_1, \ldots, \eta_n)$ a vector the components of which have an assigned covariance matrix $\mathfrak{M}$, we can briefly express the procedure for obtaining the vector $\bar{\eta}$ by means of the relation

$$\bar{\eta}=A\bar{\xi}.$$

We observe that the generation of a set of normal quantities with nonzero expectations $a_1, a_2, \ldots, a_n$ is accomplished by means of the addition $\eta_1+a_1, \ldots, \eta_n+a_n$.

In problems of detection theory (see Chapter V) normally distributed quantities serve to simulate the components of voltage vectors. When the number of elements in the message is equal to n, $2n$-dimensional distributions must be simulated. It is possible in some cases to restrict ourselves to two independent groups of quantities, each of which is n-dimensional and represents a corresponding set of components.

The voltage amplitudes are represented as the moduli of two-dimensional normal vectors. Let us consider methods of simulating multidimensional amplitude distributions when there is correlation between the voltage vectors.

We shall simulate quantities assigned by the density (5.12) corresponding to the case in which independent noise voltages are combined with a signal varying from signal element to signal element by a constant factor

$$p(x_1, \ldots, x_n; q)=$$

$$\prod_{k=1}^{n} x_k e^{-\frac{x_k^2}{2}} \int_0^{\infty} \frac{s}{q^2} e^{-\left(\frac{1}{q^2}+\sum a_k^2\right)\frac{s^2}{2}} \prod_{j=1}^{n} I_0(x_j a_j s)\, ds.$$

(We assume the noise intensity to be the same in each of the elements and we take it as unity).

The first of the simulation methods suggested reproduces essentially the derivation of this density given in Section 1, Chapter V: n Rice-distributed quantities $\eta_1, \eta_2, \ldots, \eta_n$ are generated, the kth of which has density

$$p_k(x)=xe^{-\frac{x^2+a_k^2 s^2}{2}} I_0(a_k s x),$$

where the parameter s, common to all quantities, is random and is distributed according to the Rayleigh law

$$p(s)=\frac{s}{q^2} e^{-\frac{s^2}{2q^2}}.$$

Such a procedure ensures that the process has the desired structure: the useful signal remains constant within a given sample and fluctuates from experiment to experiment; the successive noise voltages combined with it are independent.

Another simulation method operates with the voltage-vector components: having obtained $2n$ normal quantities with zero expectation and unit variance, we distribute them in pairs

$$\xi_1,\ \eta_1;\ \xi_2,\ \eta_2;\ \ldots;\ \xi_n;\ \eta_n,$$

representing the components of successive noise-voltage vectors. Then we generate two more normal numbers s_1 and s_2, again having zero expectation, but with variance equal to q^2. These numbers simulate the components of the vector representing the useful signal. By denoting by a_k the factor that distinguishes the signals in a given element from those in another element, we write the expressions for the amplitudes of signal and noise combined

$$r_k = \sqrt{(\xi_k + a_k s_1)^2 + (\eta_k + a_k s_2)^2}, \qquad k = 1, \ldots, n.$$

We observe that in the problem considered the set of $2n$ components of the signal-plus-noise voltages forms a multidimensional normal quantity with nondiagonal covariance matrix. A set having density (5.15) can be simulated by a similar procedure.

The quality of random numbers obtained on electronic computers can be assessed by means of a series of statistical criteria. Some such criteria for uniform quantities have been given in Chapter VI. We shall consider methods of checking the quality of models of multidimensional quantities.

For normal vectors there is no need for a special investigation of the goodness of fit of experimental with theoretical distributions. In fact, if the initial independent normal numbers (from which the required correlated normal numbers are generated by means of a linear transformation) have undergone goodness-of-fit tests, then their linear combinations, being sums of independent quantities, are bound to satisfy these criteria. Attention must be devoted however to determining the covariance matrix of the realized normal vector, since even a weak correlation of the initial uniform pseudo-random numbers may appreciably affect the correlation of the components of this vector. The closeness of the theoretical to the sample covariance matrix is estimated on the basis of the usual statistical rules. Thus the simplest test of the quality of a model of a multidimensional normal quantity consists in evaluating the sample second moments and comparing them with the assigned values.

A more accurate method of testing a model of a normal vector consists in evaluating the probability that the point corresponding to the components will be found within the ellipse defined by the inequality

$$X\mathfrak{M}^{-1}X \leqslant C$$

or, more explicitly,

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j \leqslant C,$$

where a_{ij} are the elements of the matrix $\mathfrak{M}^{-1}$, the inverse of the covariance matrix. In terms of the χ-distribution we can easily give an exact expression which is compared with the experimental value. The latter is determined in the following manner: having determined the n components of the multidimensional quantity to be simulated, we evaluate

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j,$$

and if the value of this quadratic form does not exceed C, we consider the experiment to be successful and add unity to the counter of the number ν of successful outcomes. In the contrary case the experiment is considered unsuccessful. Having carried out such an experiment a sufficiently large number of times N (of the order of several thousands), we find the required value of the relative frequency as ν/N. This method enables us to verify simultaneously both the normality of each component of a pseudorandom vector and the existence of the assigned correlation. Its disadvantage is that a large part of the computer memory is occupied by the matrix $\mathfrak{M}^{-1}=\|a_{ij}\|\left(\frac{n(n+1)}{2}\text{ locations}\right)$ and by the triangular transformation matrix S (the same number of locations), and this sets a limit to the number of dimensions of the quantities which may be tested.

Another, simpler method of testing the quality of a model of a normal vector $\xi=(\eta_1, \ldots, \eta_n)$ is based on the fact that the quantity

$$\zeta_n = \eta_1 + \ldots + \eta_n$$

is normal with mean value 0 (if $\mathsf{M}\eta_1 = \ldots = \mathsf{M}\eta_n = 0$) and variance

$$\mathsf{D}\zeta_n = \sum_{i,j=1}^{n} m_{ij}, \qquad \|m_{ij}\| = \mathfrak{M}.$$

This method is implemented by verifying, by the ζ_n-test, the goodness of fit of the empirical distribution of the quantity χ^2 with the normal distribution of parameters $(0, \sqrt{D\zeta_n})$.

A more complicated check is that of the quality of models of multidimensional amplitude distributions that are not determined uniquely by the second moments nor even by a finite set of quantities. Models of such distributions can be tested by evaluating the frequency with which each amplitude exceeds a number of levels, as well as the frequency with which various pairs of amplitudes exceed simultaneously assigned levels. The choice of these rather than other characteristics is due to the fact that they can be compared with known theoretical values: one-dimensional probabilities (assigned by the Rayleigh distribution) and two-dimensional probabilities (assigned by Rice's two-dimensional amplitude density). The latter corresponds to the amplitudes r_1 and r_2 built up from the normal quantities $\xi_1, \xi_2, \xi_3, \xi_4$ $(r_1^2 = \xi_1^2 + \xi_2^2,\ r_2^2 = \xi_3^2 + \xi_4^2)$ with covariance matrix

$$\mathfrak{M} = \begin{Vmatrix} \sigma^2 & 0 & a & b \\ 0 & \sigma^2 & -b & a \\ a & -b & \sigma^2 & 0 \\ b & a & 0 & \sigma^2 \end{Vmatrix}.$$

Let us write $D = \det \mathfrak{M}$. Then Rice's joint amplitude density has the form

$$p(r_1, r_2) = \begin{cases} \dfrac{r_1 r_2}{\sqrt{D}} \exp\left\{ -\dfrac{\sigma^2 (r_1^2 + r_2^2)}{2\sqrt{D}} \right\} I_0\left(\sqrt{\dfrac{a^2 + b^2}{D}}\, r_1 r_2 \right); & r_1, r_2 > 0 \\ 0, \quad r_1 \cdot r_2 < 0 \quad \text{or} \quad r_1 < 0,\ r_2 < 0, \end{cases}$$

and the corresponding distribution function

$$F(x, y) = \int_0^x \int_0^y p(r_1, r_2)\, dr_1\, dr_2$$

can be easily tabulated by expanding the function I_0 in a power series.

Joint distributions of three or more dependent amplitudes are considerably more complicated, and therefore it is not expedient to calculate the frequency with which an assigned level is exceeded at the same time by two amplitudes, since the theoretical probabilities are not available.

The quality of pseudo-random numbers can also be investigated by simulating problems that admit of exact solution. An example is provided by the calculation of the detection characteristics with amplitude-square summation (see Section 2, Chapter V).

Appendices*

I. Table of Random Numbers

86515	90795	66155	66434	56558	12332	94377	57802
69186	03393	42505	99224	88955	53758	91641	18867
41686	42163	85181	38967	33181	72664	53807	00607
86522	47171	88059	89342	67248	09082	12311	90316
72587	93000	89688	78416	27589	99528	14480	50961
52452	42499	33346	83935	79130	90410	45420	77757
76773	97526	27256	66447	25731	37525	16287	66181
04825	82134	80317	75120	45904	75601	70492	10274
87113	84778	45863	24520	19976	04925	07824	76044
84754	57616	38132	64294	15218	49286	89571	42903
75593	51435	73189	64448	31276	70795	33071	96929
73244	61870	28709	38238	76208	76575	53163	58481
23974	14783	17932	66686	64254	57598	26623	91730
32373	05312	94590	22561	70177	03569	21302	17381
59598	56774	08749	43448	28484	16325	62766	31466
91682	12904	29142	65877	64517	31466	02555	52905
87653	98088	75162	97496	59297	79636	74364	16796
79429	66186	59157	95114	16021	30890	21656	93662
85444	39453	67981	49687	36801	38666	50055	11244
85739	44326	91641	40837	93030	03675	18788	91332
84637	76154	14150	07876	41899	69207	66785	87225
59575	32764	91090	66515	05498	51512	16107	52141
81305	58846	69558	41675	88898	23775	30649	86545
29835	35801	23472	22700	39976	21279	36694	85970
32795	54313	39072	16809	22148	60102	18465	87650
37837	12507	54594	30814	23277	99497	11037	63718
58394	96952	12181	11641	83373	14726	23541	25774
74543	46849	95714	70358	95873	94136	83991	77299
77338	59570	29277	82041	06923	01795	77022	17443
21157	50634	16432	44292	20030	38547	67134	95995

---

*For explanations of Tables I and II see page 279.

Table I (Continued)

46747	15614	57723	14233	26300	80126	23963	67058
65087	17420	79559	71028	11105	43860	64747	29415
31523	70181	60580	66259	45475	23232	38489	36452
31051	78489	46105	63541	51391	62321	94557	63413
82112	45793	01251	42918	98647	32045	24312	32913
16363	91017	77362	47792	26455	56339	40125	64329
48752	72467	12333	25578	38302	40607	09174	04703
21440	77078	54895	95743	12820	53855	43806	09473
08051	77898	76323	33133	97236	33847	19881	28066
25601	33322	27768	63779	07284	64203	47631	98591
46351	98006	68132	47103	39792	20469	03412	22394
10145	24503	45377	82639	95363	22033	09662	93227
22896	04821	78192	51085	27249	90973	33421	86581
58479	11756	40940	79909	59279	65440	59937	10832
04510	93838	99861	41097	55904	30897	51144	20490
32882	44475	62628	52985	97184	75243	38108	49598
86678	06118	05562	20241	08113	89133	55420	42586
45406	03247	72727	19163	05251	63010	23874	65161
66178	94899	39761	76018	44756	08177	70297	09954
00385	04954	41398	17221	17577	63470	02495	89134
32767	45035	14305	13280	96298	17838	19095	32573
92934	57660	32300	84302	04683	94781	46083	40014
17106	13031	30146	18590	52577	42083	44904	05413
34586	40984	63908	92255	53697	28837	67279	44882
62265	99652	35244	07823	28474	64579	49830	32783
27700	63524	01708	77312	78805	06778	13283	57336
70221	49927	01538	90184	21571	73597	71368	11117
28191	84929	06817	70649	33028	24007	14863	73577
65318	12664	10543	40249	39158	12202	66946	77806
78069	49921	82444	02888	97849	25016	26861	09521
41219	99848	51620	08823	98062	44173	70712	87692
10049	75413	19045	60702	85072	15636	34982	75562
34098	25973	91865	68605	14518	48420	10643	20715
06810	01522	97114	62039	69126	79248	60948	00820
81495	67384	98286	87008	07131	25669	94275	09093
48498	84032	57014	95782	77992	94837	28824	87884
61078	17571	91916	57831	50342	82640	93678	64517
13314	64748	40943	68585	84037	32216	94790	38819
83359	93058	98270	56043	36371	07726	29952	11523
92230	23434	04513	30181	36046	97874	11872	58211

Table I (Continued)

94963	74314	33907	99969	59315	52309	93342	21435
82419	74419	83348	57837	44518	20589	33640	33858
14586	63828	24470	26119	17982	10074	05881	80362
10750	62790	19246	06711	69123	94081	73021	25354
25165	19981	01034	13229	18305	04039	30668	99537
62146	49899	85596	28215	23947	90368	49677	30608
99230	16830	01616	09704	01665	67557	90142	97074
97630	46940	04507	09945	33690	39054	51933	79347
77456	93441	34102	39925	83750	57347	31686	27094
29391	50490	32937	61626	26413	65397	53297	28931
51950	93994	05504	72118	54076	13728	13948	81906
95186	45099	56658	30765	18373	40016	77970	45800
17230	88108	47478	81898	71577	58556	01777	58752
77916	79187	39413	91229	40574	04602	61162	87186
93622	40720	00552	59380	53997	27699	35557	29701
83563	78753	56688	75496	53496	84805	09304	54066
47322	59891	04764	56768	51856	89886	23608	54650
18874	27427	55350	64148	55303	67523	70045	38043
93884	48419	18324	07146	56770	60338	14268	54342
92928	78887	13397	89114	49188	52277	26665	90439
26722	26759	81964	39814	15297	91252	99719	99399
94624	59833	82636	14536	86326	82168	59947	89263
68792	53333	12249	45116	07083	47188	33755	65621
38521	45357	60728	48298	12224	07315	37971	55539
49780	58043	47618	27883	59045	71153	74145	36200
66828	60150	39264	64860	81513	39063	12496	52088
47312	89132	83816	21519	37571	23793	14528	00820
56294	31388	36579	46587	76678	59674	55405	65813
71891	10917	71396	41691	68941	90055	94387	16156
05688	32624	56130	09695	01106	10053	64291	83350
97125	51625	08070	52065	47126	73890	40346	05698
89284	50453	74467	80685	80961	30763	37246	92137
43549	40586	81516	51880	41343	66292	02121	14564
72526	79772	76663	33372	97965	32464	63394	04344
80713	11007	01014	23840	43638	37326	87030	68210
42281	05177	68910	21809	24999	86568	77188	02130
00226	67052	12793	80539	96900	29195	76460	85299
18022	89101	09949	52569	15280	21053	75883	73989
19104	89645	04823	86417	93256	74417	25270	06569
68251	26253	81542	98017	22338	24297	96442	24748

Table I (Continued)

09236	65687	99595	29732	02164	29355	67021	79078
86058	55744	01628	60154	76141	33877	73149	52993
98654	31271	93089	91432	73210	46338	40570	75134
24230	02041	22592	73570	55838	19176	33471	37223
83760	48336	89010	41179	15624	55978	91734	09731
16070	98566	23294	48608	61162	18826	10094	87310
65772	28976	54794	53750	68663	94619	16582	08993
77384	59106	95769	94443	16357	99405	60130	16686
78120	63979	39085	12164	32556	13427	29805	65147
73325	04458	25636	89350	28926	23348	98963	00053
26453	48826	50357	66819	85259	97684	70052	81361
94451	80733	51978	99816	96777	69889	64682	91635
14786	85391	28116	71989	51460	99376	65904	88517
47904	97167	87158	36103	27949	51309	59285	52608
39322	89151	73891	63876	99274	85497	15690	97412

II. Table of Normal Quantities

0.2005	1.1922	-0.0077	0.0348	1.0423	-1.8149	1.1803	0.0033
1.1609	-0.6690	-1.5893	0.5816	1.8818	0.7390	-0.2736	1.0828
0.5864	-0.9245	0.0904	1.5068	-1.1147	0.2776	0.1012	-1.3566
0.1425	-0.2863	1.2809	0.4043	0.6379	-0.4428	-2.3006	-0.6446
0.9516	-1.7708	2.8854	0.4686	1.4664	1.6852	-0.9690	-0.0831
-0.5863	0.8574	-0.5557	0.8115	-0.2676	-1.2496	-1.2125	1.3876
1.1572	0.9990	-0.1032	0.5405	-0.6022	0.0093	0.2119	-1.4647
-0.4428	-0.5564	-0.5098	-1.1929	-0.0572	-0.5061	-0.1557	-1.2384
-0.3924	1.7981	0.6141	-1.3596	1.4943	-0.4406	-0.2033	-0.1316
0.8319	0.4270	-0.8888	0.4167	-0.8513	1.1054	1.2237	-0.7003
0.9780	-0.7679	0.8960	0.5154	-0.7165	0.8563	-1.1630	1.8800
0.4083	0.1172	0.4380	0.2153	0.2471	1.2222	-0.5332	0.6769
0.2506	0.1451	-0.1072	1.5161	-0.1149	1.7168	-0.4254	-0.4328
-0.6155	-1.6600	0.6499	-1.1385	-0.0786	0.0723	-2.1401	-0.5111
2.1961	0.8373	0.8334	0.0836	0.5572	-0.8472	0.6057	-0.2271
0.5955	-0.9567	-0.4525	1.7141	0.1166	-0.3035	-0.8648	-0.2762
0.8878	1.4399	-0.1673	1.9888	1.3912	1.1409	0.1408	1.1105
1.3596	0.6264	0.7770	-0.4714	-1.8767	-0.3315	-0.4552	0.5407
0.3961	0.3172	1.3063	1.7748	-0.7831	1.1407	-1.2317	-1.5495
1.3996	-0.5412	-0.3513	0.0959	-1.2982	-0.2092	1.4988	-0.8433

Table II (Continued)

0.6103	0.2143	-1.7647	0.8872	1.2249	0.2244	1.3989	0.2175
1.0572	0.0863	-0.5204	0.1195	0.4177	-1.3384	-1.1699	-1.5144
-0.7934	1.4052	1.5732	0.0606	2.8545	0.1878	-0.0843	0.7661
0.7095	-0.8784	-0.4474	-1.6972	1.7019	-0.0541	0.7496	0.8885
0.4189	-1.1017	-0.3966	0.2764	-0.8211	-2.1006	0.3755	0.5803
0.9204	-1.1859	0.6551	-1.1844	-0.0327	2.6107	-1.5307	0.2634
1.0469	1.3231	-1.4720	-1.5188	0.1818	-0.3886	-1.1073	0.2446
0.0373	1.0704	0.6511	0.0283	1.3351	-0.0447	1.0698	1.7287
0.8325	0.3653	0.9094	-1.2804	-0.3766	0.1126	-0.7246	-0.6078
-1.0980	-0.5706	-0.9678	-0.2011	-2.6891	0.7349	-0.2401	2.1936
0.8285	-0.9246	0.2817	-1.4527	-1.8206	-0.9273	0.1009	0.4673
0.5849	-1.2114	1.9777	-0.7401	-2.1394	-0.1418	0.5652	-0.2171
-1.2694	-0.8690	-0.6200	0.8522	0.3989	-1.6043	1.3361	-0.3627
-1.8795	2.1241	-1.0324	-0.5972	-0.5367	-1.1800	1.2464	-0.9108
-1.4345	0.8420	-2.1421	0.3184	1.7609	-1.3321	-1.7729	-0.7298
-0.2127	-0.6388	0.5343	0.9895	-0.1026	0.4739	-1.5747	0.2136
0.4394	0.6535	-1.4222	0.5028	-0.9808	-0.7944	-0.1824	-1.2026
-1.7181	1.0547	1.1764	0.9064	-1.6218	0.5456	-0.2568	-0.0006
-1.1704	2.3965	-0.1719	-1.4232	0.6767	0.1161	0.5588	-0.1572
-1.4312	-1.3493	1.2580	1.5037	-0.2647	-1.2482	-0.3690	1.4047
-0.5767	-0.1932	-0.3199	-1.2330	1.1220	-0.1292	-1.9742	-0.2299
-1.7750	-1.3836	0.5768	0.8314	0.4712	-1.8326	0.0974	0.0229
0.6989	-0.9137	0.8286	-0.6841	0.2309	0.8760	-1.3196	0.8593
1.4573	-0.5137	-0.8159	1.6976	1.6808	-0.5607	1.6192	-1.5036
-1.9278	1.3322	1.7982	-0.2411	0.0470	0.7516	-1.2719	-1.1642
-0.0307	0.2432	0.1295	0.9592	1.1758	-0.2544	-0.3663	1.8976
1.9745	-0.9768	-0.6119	-2.3175	-1.5597	0.1089	-0.8989	0.4925
-0.6722	-0.8153	0.3077	-0.4988	-1.3960	-1.9540	0.1807	-0.5275
0.8308	2.5666	0.4702	0.0564	0.2810	0.0845	0.4265	0.7492
-0.9590	-0.0971	0.4413	-1.6344	0.5451	-0.4258	-0.5102	0.3929
0.3326	-0.7979	-1.5355	-1.3690	2.0035	0.8132	0.1999	-0.3996
0.6549	0.1661	-2.4684	-0.8117	-0.1016	0.8073	-0.0469	-2.2602
-1.2105	-2.3642	-1.1261	0.1402	0.4794	-0.8293	-0.1925	-1.5751
0.5716	0.2217	0.5378	0.1267	1.3062	0.7560	1.3532	0.5975
-0.0708	1.3550	-0.8212	-0.3456	0.2366	1.1349	0.2758	0.0464
-1.1670	-0.2764	-0.9646	-0.4825	0.9841	0.9214	-0.9516	0.2055
-1.7540	1.4139	-0.6992	-0.0217	-0.8669	1.3461	0.3107	-1.6222
-0.1531	1.3783	-0.0558	0.5885	-1.1431	-1.5575	-0.0893	0.9620
0.0814	-0.6202	-1.5163	-0.5086	0.3290	-2.6179	1.3521	0.9347
1.0686	0.4729	0.0907	0.6189	2.1538	-1.5846	0.0904	-0.8897

Table II (Continued)

-0.9519	2.4481	-1.2937	-0.3668	0.3879	-0.4519	-0.7694	1.4488
-1.2364	-0.4280	-0.5502	-1.1491	0.0441	-0.1506	0.5166	0.4555
0.2610	0.7267	1.0333	0.4426	-0.5912	-0.6600	-1.3500	-1.1604
-1.1613	-2.1545	-0.1767	-0.3837	0.2718	1.0682	0.7272	-1.9370
0.6500	0.8222	1.7626	-0.0396	-1.4831	0.7821	0.7722	-0.1627
1.6165	-0.7251	-0.9607	1.1737	1.9487	1.3683	0.1228	1.7110
0.0475	-0.0944	0.6800	0.2583	-1.3042	-0.5009	1.5905	0.2739
-1.4437	0.9985	-0.2006	1.5134	-0.1220	-1.2408	0.9547	0.9419
0.8907	0.4311	0.4879	-0.5611	-0.4382	-0.0232	0.7202	-1.6899
-1.1346	-0.8556	-1.5648	-1.4676	-0.5550	1.9783	-0.6002	-0.8623
1.4385	-0.4879	-0.0615	3.3023	0.0828	-0.4150	-0.3619	-1.1449
0.2328	0.2764	0.5612	1.2642	0.0228	-0.0113	-1.1525	0.8575
0.2062	0.5703	-0.7782	-0.3347	0.6958	-1.6643	-0.1400	-0.4758
-1.5345	0.2521	0.0286	-1.0374	-0.3167	0.0337	-1.5653	-0.3404
-0.4928	0.6253	-2.3459	-0.6131	-0.7769	-0.8436	0.0709	1.6279
-0.5841	2.6767	1.7088	-0.6897	0.4119	0.7105	1.5529	-0.8765
0.1518	-0.7753	-1.4071	-0.3088	-0.7625	1.2727	-1.1864	0.5924
0.3180	0.0787	-0.9492	0.8690	-0.2502	-0.2786	-0.1053	0.9414
1.1551	-0.1413	-1.9574	0.8299	0.0078	0.3747	0.4133	-0.0213
0.1820	-0.7459	0.2789	-0.4731	-0.8643	1.1979	0.6021	0.1031
-0.4705	1.8856	-1.3087	-0.4818	0.0037	-0.1947	0.5400	0.2916
0.9670	0.6316	0.9936	-0.3840	0.0104	-1.8476	1.0431	-0.7333
-1.4820	0.2212	1.0139	1.6695	0.6186	1.0103	-0.1774	0.7496
1.1444	1.5919	-0.4227	-1.0081	-0.3190	-1.4802	-1.0290	1.2259
0.0293	-1.4862	-1.5943	0.4494	1.6794	1.4924	0.3306	-0.6443
0.3489	1.2612	1.6360	1.1331	0.7388	0.2482	-0.9710	-0.0743
-0.7011	1.4512	-0.1605	1.2953	0.4469	2.6422	-0.5301	-0.3351
0.9104	-0.0423	-0.6756	0.0033	-0.8384	0.1191	-0.9857	0.0763
1.4756	0.4546	-0.7222	-0.5631	0.4012	-0.4609	-0.3180	-0.6076
1.3478	2.3351	0.1138	-0.1341	1.3255	0.1626	0.4069	0.4443
-0.5155	0.8359	0.6713	0.4304	0.2434	-0.4370	1.9007	2.3921
0.4917	0.8378	0.2625	-0.6391	0.4374	0.4508	2.0110	1.0462
1.5164	-0.7221	-0.6674	-0.6961	-0.4910	0.9222	0.1287	-0.3559
-0.4471	0.2909	0.0715	1.3707	-1.6617	-0.9242	0.4884	0.8732
0.9266	-0.4778	0.7206	0.7800	0.0541	-0.9749	-0.3148	-1.7382
1.1780	-1.6038	0.3676	0.2777	-0.6000	-0.3385	-0.1781	0.0006
-0.8993	0.1293	0.3592	-0.6395	0.2762	0.2523	-0.4792	-1.9707
0.6425	0.1769	1.1089	1.0447	1.6418	1.3131	-0.4279	-0.2146
0.4918	-0.7745	0.5588	-0.2296	0.7240	-0.7744	1.0531	-0.3648
0.6104	-0.7964	-1.0848	1.0418	-2.3557	-2.2113	-0.2154	-0.4277

Table II (Continued)

0.4524	-0.3759	-1.1044	-0.7922	-0.3908	0.5027	-0.8777	0.8350
1.3386	-0.9384	0.9778	0.5967	0.0569	-0.4893	-0.1772	-0.1316
0.3578	-0.0375	-0.2984	-0.2797	-1.2492	0.4011	-2.9682	-0.3262
-0.0892	1.5757	0.9775	-0.6173	2.0789	-0.5821	0.4185	-1.1778
-0.7167	-1.9664	-2.8695	-0.7468	0.1990	-0.2504	-0.6488	-0.6664
-0.9094	-0.3657	0.0288	-0.4141	1.7280	1.7529	1.2845	-2.5891
-2.1042	-0.3644	-0.0820	0.1336	0.2047	0.6016	0.0180	1.6723
-1.2015	-0.4216	1.2343	0.7220	-1.0582	-1.6844	1.3734	2.2182
-1.1590	1.3998	-0.7100	0.4390	0.3926	0.0069	-0.9308	0.4610
-0.1055	-0.6711	-0.5184	0.4196	-0.7284	0.3841	0.2318	0.4297
-0.4567	1.5014	2.2306	0.1521	-1.3206	0.2484	-1.0647	1.3832
0.8239	0.5671	-0.8463	-0.8796	-0.4575	0.8009	0.1745	0.8412
1.6095	-1.4104	0.9356	-0.5948	-1.4605	0.3091	-0.7754	-0.3175
-1.8891	-2.4284	-0.3966	-0.0619	0.8885	0.2336	-0.6679	-0.9313
0.2825	0.1674	-0.5899	-0.0527	-0.6281	1.8808	0.3292	-0.5599
-1.2549	1.8023	-0.4091	0.7098	-1.1160	0.3299	-1.3397	-0.4159
0.6437	1.2778	0.8326	-0.5625	0.9745	0.9824	-0.0795	0.8991
0.8561	-0.2334	2.1016	0.1617	-0.0937	0.7040	-2.1222	0.6033
-0.7138	1.7661	0.3209	-1.2491	-0.1967	-0.8866	0.1061	-0.0075
-0.3168	-0.2093	-0.0818	0.4804	0.6276	-0.5237	2.0092	-2.1167
-0.5841	1.0182	-0.1987	1.0343	1.1264	1.7410	-1.2655	-0.5511
-0.0420	-0.1988	1.3127	1.5628	2.3005	2.8135	0.4605	0.2354
0.4347	0.4712	-0.6811	1.8966	-0.9681	1.8779	0.2103	0.9124
0.0955	1.0883	1.0695	-1.6493	1.2889	-0.6178	0.9129	-0.1627
-0.8127	0.2933	0.8815	0.9142	1.2547	1.6200	-0.1370	0.0929

NOTES

Random digits simulate the results of a sequence of identical experiments, in each of which a digit is extracted "by lot" from the set 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 (that is, the probability of occurrence of each of the digits is equal to 1/10). The experiments are independent of each other.

The 5000 digits given in Table I (page 273) are the first decimal digits of pseudo-random numbers $\{\gamma\}$ used on the "Strela" computer. A description and the results of tests for "randomness" carried out on these numbers are found in the paper [57]. The digits are divided for the sake of convenience into groups of five.

Normal quantities simulate a sequence of values of a normally distributed (Gaussian) random quantity with mean 0 and variance 1.

The 1000 numbers given in Table II (page 276) have been calculated on a "Strela" computer from pseudo-random numbers $\{\gamma\}$ according to a method suggested in the paper [7] using the formula

$$\zeta = 0.01\zeta_5\left(97 + \zeta_5^2\right),$$

where

$$\zeta_5 = \sqrt{0.6}\sum_{i=1}^{5}(2\gamma_i - 1)$$

(each ζ is calculated from five values of γ).

References

1. A. F. AKKERMAN and D. K. KAIPOV, 1960. Application of the Monte Carlo method to calculations of the passage of γ-quanta through matter, Trudy In-ta yadernoi fiziki AN KazSSR, Vol. 3, 106-114.
2. Atlas of Neutral Cross Sections, 1959. Atomizdat, 2nd edition.
3. N. S. BAKHVALOV, 1959. On the approximate evaluation of multiple integrals, Vestnik Mosk. Gos. Univ., ser. mat., mekh., astr., fiz., No. 4, 3-18.
4. G. P. BASHARIN, 1959. On the statistical estimation of the entropy of a sequence of independent random quantities, Teoriya veroyatn. i ee primen., Vol. 4, No. 3.
5. Yu. A. BLAGOVESHCHENSKII, 1962. On an effective Monte Carlo method for the solution of certain problems, Voprosy teorii matemat. mashin, Collection 2.
6. L. N. BOL'SHEV, 1960. Probability estimates, Teor. veroyatn. i ee primen., Vol. 5, No. 4.
7. L. N. BOL'SHEV, 1959. Transformations of random quantities, Teor. veroyatn. i ee primen., Vol. 4, No. 2, 136-149.
8. L. F. BORODIN, 1960. Ideal identification equipment for complex signals, Radiotekhnika, Vol. 15, No. 8.
9. G. W. BROWN, 1958. The Monte Carlo method, in Collection: Modern Mathematics for Engineers, edited by E. Beckenbach (transl.), In. Lit.
10. N. P. BUSLENKO, 1961. The solution of problems of queueing theory by the simulation method on electronic digital computers, Problemy peredachi informatsii, Collection 9.
11. N. P. BUSLENKO and Yu. A. SHREIDER, 1961. The Method of Statistical Experiments (the Monte Carlo Method) and Its Practical Realization In Digital Computers, Fizmatgiz.
12. A. WALD, 1960. Sequential Analysis (transl.), Fizmatgiz.
13. B. L. VAN DER WERDEN, 1960. Mathematical Statistics (transl.) In. Lit.
14. A. A. VARFOLOMEEV, D. I. GOLENKO and G. A. SVETLOLOBOV, 1958. Characteristics of electromagnetic cascades in photo-emulsion allowing for the effect of the medium on radiation processes, Dokl. Akad. Nauk SSSR, Vol. 122, No. 5, 785-787.

15. E. S. VENTTSEL', 1962. Probability Theory, Fizmatgiz, 2nd edition.
16. V. S. VLADIMIROV, 1956. Use of the Monte Carlo method for finding the least characteristic number and the corresponding eigen-function of a linear integral equation, Teor. veroyatn. i ee primen., Vol. 1, No. 1, 113-130.
17. V. S. VLADIMIROV, 1960. The approximate evaluation of Wiener integrals, Usp. mat. nauk, Vol. 15, No. 4, 129-135.
18. V. S. VLADIMIROV and I. M. SOBOL', 1958. Calculation of the least eigen-value of Peierls' equation by the Monte Carlo method, Vychislit. matematika, No. 3, 130-137.
19. I. M. GEL'FAND, S. M. FEINBERG, A. S. FROLOV and N. N. CHENTSOV, 1958. Application of the Monte Carlo method to the solution of the kinetic equation, Proceedings of the Second International Geneva Conference on the Peaceful Uses of Atomic Energy, No. 2141.
20. I. M. GEL'FAND, A. S. FROLOV and N. N. CHENTSOV, 1958. The evaluation of Wiener integrals by the Monte Carlo method, Izv. vys. uch. zav. SSSR, ser. matem., No. 5, 32-45.
21. I. M. GEL'FAND and N. N. CHENTSOV, 1956. Numerical evaluation of Wiener integrals, Zh. eksp. teoret. fiz., Vol. 31, No. 6, 1106-1107.
22. I. M. GEL'FAND and A. M. YAGLOM, 1956. Integration in functional spaces and its application in quantum physics, Usp. mat. nauk, Vol. 11, No. 1, 77-114.
23. B. V. GNEDENKO, 1961. Course of Probablity Theory, Fizmatgiz, 3rd edition.
24. D. I. GOLENKO, 1959. The generation of random quantities with arbitrary distribution law, Vychislit. matematika, No. 5, 83-92.
25. D. I. GOLENKO, 1959. Calculation of the characteristics of certain stochastic processes by the Monte Carlo method, Vychislit. matematika, No. 5, 93-108.
26. D. I. GOLENKO, V. P. SMIRYAGIN, V. Ya. KAPLANSKII and Yu. M. SHIVALIN, 1960. Random-number generator for the "Strela" electronic computer, Computing Center AN SSSR.
27. D. I. GOLENKO, 1960. Some problems of the calculation of stochastic processes by the Monte Carlo method, Proceedings of the All-Union Conference "New Developments of Computational Mathematics and Computing Engineering," Kiev.
28. D. I. GOLENKO and V. P. SMIRYAGIN, 1960. Random-number generators, Proceedings of the Mathematical Institute of the Hungarian Academy of Sciences, series A, Vol. 15, No. 3.
29. V. B. DAVENPORT and V. L. ROOT, 1960. Introduction to the Theory of Random Processes and Noise (transl.), In. Lit.

30. I. A. DANIL'CHENKO, 1961. The evening-out of the probabilities of occurrence of binary digits by superimposing direct and inverse representations, Voprosy teorii matem. mashin, Collection 2.
31. A. M. DLIN, 1953. Mathematical Statistics in Engineering, Sovietskaya Nauka Editions.
32. J. L. DOOB, 1956. Stochastic Processes (transl.), In. Lit.
33. I. V. DUNIN-BARKOVSKII and N. V. SMIRNOV, 1955. Probability Theory and Mathematical Statistics in Engineering (General Part), Gostekhizdat.
34. B. DAVISON, 1960. Theory of Neutron Transfer, Atomizdat.
35. S. M. ERMAKOV and V. G. ZOLOTUKHIN, 1960. Polynomial approximations and Monte Carlo method, Teor. veroyatn. i ee primen., Vol. 4, No. 4, 473-476.
36. T. N. ZUBAREV, 1959. Introduction to the Theory of Nuclear Reactors, MAI Editions.
37. M. KADYROV, 1936. Tables of Random Numbers, Izd. Sredneaziatskogo Gos. un-ta, Tashkent.
38. D. KERTISS, 1957. Monte Carlo methods for the iteration of linear operators, Usp. mat. nauk, Vol. 12, No. 1, 149-174.
39. A. I. KITOV and N. A. KRINITSKII, 1961. Electronic Digital Computers and Programming, Fizmatgiz., 2nd edition.
40. N. M. KOROBOV, 1957. Approximate evaluation of multiple integrals by means of methods of number theory, Dokl. Akad. Nauk SSSR, Vol. 115, No. 6, 1062-1065.
41. N. M. KOROBOV, 1959. Approximate evaluation of multiple integrals, Vestnik Mosk. Gos. Univ., No. 4.
42. N. M. KOROBOV, 1950. On certain problems of the uniform distribution, Izv. Akad. Nauk SSSR, ser. matem., Vol. 14, 215-238.
43. H. CRAMER, 1948. Mathematical Methods of Statistics (transl.), In. Lit.
44. G. I. MARCHUK, 1958. Numerical Methods of Calculation of Nuclear Reactors, Atomizdat.
45. I. MATYASH and Ya. SHILKHANEK, 1960. Generator of random processes with assigned matrix of the spectral densities, Avtomatika i telemekhanika, Vol. 21, No. 1, 29-35.
46. V. E. MILNE, 1955. Numerical Solution of Differential Equations (transl.), In. Lit.
47. K. MORTON, Evaluation of the probability of escaping resonant capture by the Monte Carlo method, Proceedings of the Second International Geneva Conference on the Peaceful Uses of Atomic Energy, Collected Reports of Foreign Scientists, Vol. 2, Neutron Physics, 594-599.

48a. A. G. POSTNIKOV, 1960. Arithmetical Simulation of Random Processes, Trudy Matem. in-ta im. V. A. Steklova, Vol. 57, Izd-vo Akad. Nauk SSSR.

48b. Reception of Signals in the Presence of Noise, 1960. Collected articles edited by L. S. Gutkin, In. Lit.

49. V. S. PUGACHEV, 1957. Theory of Random Functions and Its Application to Automatic Control Problems, Gostekhizdat.

50. G. K. RAKOV, 1958. Generation of a random quantity on high-speed computers, Automatic Control and Computer Engineering (collected articles), GITI editions, Moscow-Leningrad.

51. B. I. RAMEEV and Yu. A. SHREIDER, 1958. Solution of the direct problem of the theory of resistance logging on special-purpose computers, Vopr. teor. matem. mashin, Collection 1, 172-180.

52. T. A. SARYMSAKOV, 1954. The Bases of the Theory of Markov Processes, Gostekhizdat.

53. B. A. SEVAST'YANOV, 1951. The theory of branched random processes, Usp. mat. nauk, Vol. 6, No. 6 (46).

54. Yu. B. SINDLER, 1958. The use of digital computers for the solution of problems of microwave radio-link communication by the method of statistical trials, Nauch. Dokl. Vyshch. Shkol., Radiotekhnika i elektronika, No. 1, 81-85.

55. Yu. B. SINDLER, 1960. Approximate calculation and simulation of noise buildup in radio links, Sbornik trudov nauchno-tekhnich. ob-va im. A. S. Popova, No. 2, 227-255.

56. I. M. SOBOL', 1957. Multidimensional integrals and Monte Carlo method, Dokl. Akad. Nauk SSSR, Vol. 114, No. 4, 706-709.

57. I. M. SOBOL', 1958. Pseudo-random numbers for the "Strela" computer, Teor. veroyatn. i ee primen., Vol. 3, No. 2, 205-211.

58. I. M. SOBOL', 1961. Accurate estimates of the error of multidimensional quadrature formulas for functions of the W_1 and H_1 classes. Zh. vychisl. mat. i matematich. fiz., Vol. 1, No. 2.

59. I. M. SOBOL', 1960. On the solution of Peierls' equation by the Monte Carlo method, Teor. veroyatn. i ee primen., Vol. 5, No. 3, 361-366.

60. I. M. SOBOL', 1961. On the evaluation of multidimensional integrals, Dokl. Akad. Nauk SSSR, Vol. 139, No. 4, 821-823.

61. I. M. SOBOL', 1961. On the evaluation of infinite integrals, Zh. vychisl. mat. i matematich. fiz., Vol. 1, No. 5, 917-922.

62. I. Z. FISHER, 1959. Application of the Monte Carlo method in statistical physics, Usp. fiz. nauk, Vol. 69, no. 3, 349-369.

63. A. S. HOUSEHOLDER, 1956. Fundamentals of Numerical Analysis (transl.), In. Lit.
64. V. V. CHAVCHANIDZE, 1955. The method of random trials (Monte Carlo method), Trudy in-ta fiziki Akad. Nauk Gruz. SSR, Vol. 3, 105.
65. V. V. CHAVCHANIDZE, 1955. Application of the method of random trials to the calculation of an internuclear cascade, Izv. Akad. Nauk SSSR, ser. fiz. Vol. 19, No. 6, 629-638.
66. V. V. CHAVCHANIDZE, R. S. SHADURI and V. A. KUMSISHVILI, 1958. Calculation by the Monte Carlo method of an electron-photon cascade in lead, Zh. eksp. teoret. fiz., Vol. 34, No. 4, 912-915.
67. Yu. A. SHREIDER, 1955. The method of statistical trials (Monte Carlo) and its use in digital computers, Priborostroenie, No. 7, 1-5.
68. Yu. A. SHREIDER, 1958. Solution of systems of linear algebraical equations by the Monte Carlo method, Voprosy teorii matemat. mashin, Collection 1, 167-171.
69. Nuclear Reactors, 1956. Part I, Physics of Nuclear Reactors (transl.), In. Lit.
70. H. AKAIKE, 1956. Monte Carlo method applied to the solution of simultaneous linear equations, Ann. Inst. Statist. Math., Tokyo, Vol. 7, 107-113.
71. H. AKAIKE, 1956. On optimum character of von Neumann's Monte Carlo model, Ann. Inst. Statist. Math., Tokyo, Vol. 7, 183-193.
72. H. AKAIKE and Y. SAIGUSA, 1957. Statistical analysis of some sequential phenomena (transl.), Tokey suri kensyuse ikho, Vol. 5, No. 1, 58-65.
73. G. E. ALBERT, 1956. A general theory of stochastic estimates, of the Neumann series for the solutions of certain Fredholm integral equations and related series, Symposium on Monte Carlo methods, 37-46.
74. H. J. ARNOLD, B. D. BUCHER, H. F. TROTTER and J. W. TUKEY, 1956. Monte Carlo techniques in a complex problem about normal samples, Symposium on Monte Carlo methods, 80-88.
75. J. BASS and J. GUILLOUD, 1958. Méthode de Monte-Carlo et suites uniformément denses, Chiffres, 1, 149-155.
76. W. F. BAUER, 1958. The Monte Carlo method, J. Soc. Industr. Appl. Math., Vol. 6, 438-451.
77. L. A. BEACH and R. B. THEUS, 1956. Stochastic calculations of gamma ray diffusion, Symposium on Monte Carlo methods, 103-122.

78. M. J. BERGER, 1955. Reflection and transmission of gamma radiation by barriers, Monte Carlo calculation by a collision-density method, J. Research NBS, Vol. 55, No. 6, 343-350.

79. M. J. BERGER, 1956. An application of the Monte Carlo methods to a problem in gamma ray diffusion, Symposium on Monte Carlo methods, 89-102.

80. M. J. BERGER and J. W. COOPER, 1959. Reflection of fast neutrons from water, J. Research NBS, Vol. 63A, 101-144.

81. M. J. BERGER and D. J. RASO, 1960. Monte Carlo calculations of gamma-ray backscattering, Radiation Research, Vol. 12, 20-37.

82. E. BOFINGER and V. J. BOFINGER, 1958. A periodic property of pseudo-random sequences, J. Assoc. Comput. Machinery, Vol. 5, 261-265.

83. L. BOUCKAERT, 1956. Les méthodes de Monte-Carlo, Rev. questions sci., Vol. 17, 344-359.

84. G. BOUQUET, M. FEIX, P. NICOURD and C. SAJALOLI, 1958. Emploi de la méthode de Monte Carlo pour la détermination du volume critique d'un cylindre, C. R. Acad. Sci., Vol. 246, No. 9, 1382-1384.

85. S. H. BROOKS, 1958. A discussion of random methods for seeking maxima, Operat. Res., Vol. 6, No. 2, 244-251.

86. G. W. BROWN, 1951. History of RAND's random digits, NBS Appl. Math. Series, No. 12, 31-32.

87. J. W. BUTLER, 1956. Machine sampling from given probability distributions, Symposium on Monte Carlo methods, 249-264.

88. L. L. CAMPBELL, 1959. Two properties of pseudo-random sequences, IRE Trans., Inform. Theory, Vol. 5, No. 1, 32.

89. E. CANSADO, 1958. Sobre la inversion de matrices de Leontief, Trab. estadist., Vol. 9, No. 3, 203-264.

90. E. D. CASHWELL and C. J. EVERETT, 1959. A Practical Manual on the Monte Carlo Method for Random Walk Problems, Pergamon Press.

91. J. E. CERTAINE, 1958. On sequences of pseudo-random numbers of maximal length, J. Assoc. Comput. Machinery, Vol. 5, 353.

92. F. CERULUS and R. HAGEDORN, 1958. A Monte Carlo method to calculate multiple phase space integrals, I, Nuovo Cimento, Vol. 9, Suppl., No. 2, 646-658.

93. F. CERULUS and R. HAGEDORN, 1958. A Monte Carlo method to calculate multiple phase space integrals, II, Nuovo Cimento, Vol. 9, Suppl., No. 2, 659-677.

94. J. CHARLES and P. VANDENPLAS, 1959. La méthode de Monte-Carlo, I, Schémas stochastiques associés à un problème de mesure à données incomplètes, Bull. Assoc. ingrs. issue Ecole applic. artill. et génie, Vol. 37, No. 2, 1-6.
95. J. M. COOK, 1957. Rational formulas for the production of a spherically symmetric probability distribution, MTAC, Vol. II, No. 58, 81-82.
96. R. R. COVEYOU, 1960. Serial correlation in the generation of pseudo-random numbers, J. Assoc. Comput. Machinery, Vol. 7, No. 1, 72-74.
97. J. E. CREW, Y. D. HILL and L. S. LAVATELLI, 1957. Monte Carlo calculation of single pion production by pions, Phys. Rev., Vol. 106, 1051-1056.
98. I. CRONE, 1955. Einige Anwendungsmöglichkeiten der Monte-Carlo-Methode. Fortschritte der Physik, Vol. 3, No. 3, 97-132.
99. P. CUER and J. COMBE, 1954. Sur la réalisation d'un technique de Monte-Carlo pour étudier le passage de nucléons de grande énergie à travers les noyaux, C. R. Acad. Sci., Vol. 238, 1799-1801.
100. J. H. CURTISS, 1949. Sampling methods applied to differential and difference equations, Proc. seminar on sci. comput., IBM, New York.
101. J. H. CURTISS, 1956. A theoretical comparison of the efficiencies of two classical and a Monte Carlo method for computing one component of the solution of a set of linear albegraic equations, Symposium on Monte Carlo methods, 191-233.
102. R. E. CUTKOSKY, 1951. A Monte Carlo method for solving a class of integral equations. J. Research NBS, Vol. 47, No. 2, 113-115.
103. G. DAHLQUIST, 1954. Monte-Carlo-metoden, Nord. Mat. tidskr., Vol. 2, 27-43.
104. D. H. DAVIS, 1960. Monte Carlo calculation of molecular flow rates through a cylindrical elbow and pipes of other shapes. J. Appl. Phys., Vol. 31, 1169-1176.
105. P. DAVIS and P. RABINOWITZ, 1956. Some Monte Carlo experiments in computing multiple integrals, MTAC, Vol. 10, No. 53, 1-8.
106. N. D. DIAMANTIDES, 1956. Analogue computer generation of probability distributions for operations research, Commun. and Electronics, No. 23, 86-91.
107. R. P. DILWORTH and E. ACKERLIND, 1957. The analysis of post detection integration systems by Monte Carlo methods, JRE Nat. Convent. Rec., Vol. 5, No. 2, 40-47.

108. N. DISMUKE, 1956. Monte Carlo computations, Symposium on Monte Carlo methods, 52-62.

109. E. L. DODD, 1942. Certain tests for randomness applied to data grouped into small sets, Econometrica, Vol. 10, 249-257.

110. E. L. DOOD, 1942. A transformation of Tippett random sampling numbers into numbers normally distributed, Boletino matemat, Vol. 15, 73-77.

111. M. D. DONSKER and M. KAC, 1950. A sampling method for determining the lowest eigenvalue and the principal eigenfunction of Schrödinger's equation. J. Research NBS, Vol. 44, 551-557.

112. J. DOSTROVSKY, P. RABINOWTIZ and R. BIVINS, 1958. Monte Carlo calculations of nuclear evaporation processes, I, Systematics of nuclear evaporation, Phys. Rev., Vol. 111, No. 6, 1659-1676.

113. J. DOSTROVSKY, Z. FRAENKEL and P. RABINOWITZ, 1958. A Monte-Carlo calculation of fission-spallation competition, Proceedings of the Second International Geneva Conference on the Peaceful Uses of Atomic Energy, No. 1615, 1-27.

114. J. DOSTROVSKY, Z. FRAENKEL and G. FRIEDLANDER, 1959. Monte Carlo calculations of nuclear evaporation processes, III, Applications to low-energy reactions, Phys. Rev., Vol. 116, No. 3, 683-702.

115. J. DOSTROVSKY, Z. FRAENKEL and L. WINSBERG, 1960. Monte Carlo calculations of nuclear evaporation processes, IV, Spectra of neutrons and charged particles from nuclear reactions, Phys. Rev., Vol. 118, No. 3, 781-791.

116. J. DOSTROVSKY, Z. FRAENKEL and P. RABINOWITZ, 1960. Monte Carlo calculations of nuclear evaporation processes, V, Emission of particles heavier than He^4. Phys. Rev., No. Vol. 118, No. 3, 791-793.

117. V. DUPAČ, 1956. Stochastické početni methody, Čas. pro pešt. mat., Vol. 81, 55-68.

118. D. DURAND and J. A. GREENWOOD, 1957. Random unit vectors, II, Usefulness of Gram-Charlier and related series in approximating distributions, Ann. Math. Statistics, Vol. 28, 978-986.

119. H. P. EDMUNDSON, 1953. Monte Carlo matrix inversion and recurrent events, MTAC, Vol. 7, No. 41, 18-21.

120. L. W. EHRILICH, 1959. Monte Carlo solutions of boundary value problems involving the difference analogue of

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + Ky^{-1}\frac{\partial u}{\partial y} = 0.$$

J. Assoc. Comput. Machinery, 1959, Vol. 6, No. 2, 204-218.

121. E. ENGELER, 1958. Über die Monte-Carlo-Methode, Verein. Schweiz. Versich.—Math., Vol. 58, 67-76.
122. W. FICKETT and W. W. WOOD, 1960. Shock Hugoniots for liquid argon, Phys. Fluids, Vol. 3, 204-209.
123. G. E. FORSYTHE, 1951. Generation and testing of random digits at the National Bureau of Standards, Los Angeles, NBS, Appl. Math. Series, No. 12, 34-35.
124. G. E. FORSYTHE, 1951. Generation and testing of 1217370 random binary digits on SWAC, Bull. Amer. Math. Soc., Vol. 57, 304.
125. G. E. FORSYTHE and R. A. LEIBLER, 1950. Matrix inversion by a Monte Carlo method, MTAC, Vol. 4, 127-129.
126. R. FORTET, 1952. On the estimation of an eigenvalue by an additive functional of a stochastic process, with special reference to the Kac-Donsker method, J. Research NBS, Vol. 48, No. 1, 68-75.
127. E. FRANCKX, 1956. La méthode de Monte-Carlo, Assoc. actuair. Belges, No. 58, 89-101.
128. E. FRANCKX, 1959. La méthode de Monte Carlo, II, Remarques sur une problème de Monte-Carlo, Bull. Assoc. ingrs. issus Ecole applic. artill. et génie, Vol. 37, No. 2, 7-9.
129. J. FRANKLIN, 1958. On the equidistribution of pseudorandom numbers, Quart. Appl. Math., Vol. 16, No. 2, 183-188.
130. A. S. FRASER, 1958. Monte Carlo analyses of genetic models, Nature, Vol. 181, No. 4603, 208-209.
131. J. GANI and P. A. P. MORAN, 1955. The solution of dam equations by Monte Carlo methods, Austral. J. Appl. Sci., Vol. 6, 267-273.
132. G. GOERTZEL and M. H. KALOS, 1958. Monte Carlo methods in transport problems, Progress in nuclear energy, ser. 1, Phys. and math., Vol. 2, 315-369, Pergamon Press.
133. M. L. GOLDBERGER, 1948. The interaction of high-energy neutrons and heavy nuclei, Phys. Rev., Vol. 74, No. 10, 1269-1277.
134. I. J. GOOD, 1953. The serial test for sampling numbers and other tests for randomness. Proc. Cambridge Philos. Soc., Vol. 49, No. 2, 276-284.
135. W. GOOD and R. JOHNSTON, 1959. A Monte Carlo method for criticality problems, Nuclear Sci. and Eng., Vol. 5, 371-375.
136. B. F. GREEN, I. E. K. SMITH and L. KLEM, 1959. Empirical tests of an additive random number generator, J. Assoc. Comput. Machinery, Vol. 6, No. 4, 527-537.
137. R. E. GREENWOOD, 1955. Coupon collector's test for random digits, MTAC, Vol. 9, No. 49, 1-5.

138. F. GRUENBERGER, 1950. Tests of random digits, MTAC, Vol. 4, No. 32, 244-245.

139. F. GRUENBERGER and A. M. MARK, 1951. The d^2 test of random digits, MTAC, Vol. 5, No. 34, 109-110.

140. A. HAITSMA, 1959. Monte-Carlo en nabootsing, Statist. neerl., Vol. 13, No. 3, 395-406.

141. J. H. HALTON, 1960. On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, Numerische Math., Vol. 2, No. 1-2, 163-171.

142. J. H. HALTON and D. C. HANDSCOMB, 1957. A method for increasing the efficiency of Monte Carlo integrations, J. Assoc. Comput. Machinery, Vol. 4, No. 3, 329-340.

143. H. C. HAMAKER, 1949. A simple technique for producing random sampling numbers, Proc. Nederl. Akad. Wetensch., Vol. 52, 145-150.

144. P. C. HAMMER, 1951. Calculation of shielding properties of water for high-energy neutrons, NBS Appl. Math. Series, No. 12, 21-23.

145. P. C. HAMMER, 1951. The mid-square method of generating digits, NBS Appl. Math. Series, No. 12, 33.

146. J. M. HAMMERSLEY, 1956. Conditional Monte Carlo, J. Assoc. Comput. Machinery, Vol. 3, No. 2, 73-76.

147. J. M. HAMMERSLEY and J. G. MAULDON, 1956. General principles of antithetic variates, Proc. Cambridge Philos. Soc., Vol. 52, No. 3, 476-481.

148. J. M. HAMMERSLEY and K. W. MORTON, 1954. Poor man's Monte-Carlo, J. Royal Statist. Soc., ser. B., Vol. 16, No. 1, 23-38.

149. J. M. HAMMERSLEY and K. W. MORTON, 1956. A new Monte Carlo technique: antithetic variates, Proc. Cambridge Philos. Soc., Vol. 52, 449-475.

150. J. M. HAMMERSLEY and J. A. NELDER, 1955. Sampling from an isotropic Gaussian process. Proc. Cambridge Philos. Soc., Vol. 51, 652-662.

151. J. HARDY, G. G. SMITH and D. KLEIN, 1960. Experimental check of a Monte Carlo-calculated distribution of resonance neutron capture in a gold rod, Nuclear Sci. and Eng., Vol. 7, 263-267.

152. J. HARLING, 1958. Simulation techniques in operational research, Operat. Research. Quart., Vol. 9, No. 1, 9-21.

153. H. HARMUTH, 1954. Ein Statistisches Verfahren zur Lösung der Laplaceschen Differentialgleichung durch elektronische Rechenmaschinen, Acta phys. Austriaca, Vol. 9, No. 1, 27-32.

154. R. K. HAYWARD, E. L. BUBB and H. W. FENSOM, 1957. Computer selects premium bond winners, Electronics, Vol. 30, No. 7, 138-143.
155. J. HEINHOLD, 1956. Rechenautomaten und matematische Statistik, II, MTW-Mitt., Vol. 3, No. 6, 265-279.
156. J. S. HICKS and R. F. WHEELING, 1959. An efficient method for generating uniformly distributed points on the surface of an n-dimensional sphere, Communs. Assoc. Comput. Mach., Vol. 2, No. 4, 17-19.
157. S. HOERNER, 1957. Herstellung von Zufallszahlen auf Rechenautomaten, ZAMP, Vol. 8, No. 1, 1-88.
158. J. HOFFMAN, N. METROPOLIS and V. GARDINER, 1955. Study of tumor cell populations by Monte Carlo methods, Science, Vol. 122, No. 3167, 465-466.
159. H. B. HORTON, 1948. A method for obtaining random-numbers, Ann. Math. Statistics, Vol. 19, 81-85.
160. H. B. HORTON and R. T. SMITH III, 1949. A direct method for producing random digits in any number system, Ann. Math. Statistics, Vol. 20, No. 1, 82-90.
161. A.S. HOUSEHOLDER, 1951. Neutron age calculations in water, graphite and tissue, NBS Appl. Math. Series, No. 12, 6-8.
162. H. ISAAKSSON, 1958. Generator for tilfaeldige tal, Teleteknik, Vol. 9, No. 4, 175-186.
163. M. ISIDA and H. IKEDA, 1956. Random-number generator, Ann. Inst. Statist. Math., Tokyo, Vol. 8, 119-126.
164. T. ITOH and T. MUSHA, 1960. Monte Carlo calculations of the motions of electrons in helium, J. Appl. Phys., Vol. 31, 744-745.
165. N. H. JENNINGS and J. H. DICKINS, 1958. Computer simulation of peak hour operations in a bus terminal, Manag. Sci., Vol. 5, No. 1, 106-120.
166. W. N. JESSOP, 1956. Monte Carlo methods in industrial problems, Appl. Statistics, Vol. 5, 158-165.
167. D. L. JOHNSON, 1956. Generating and testing pseudo-random numbers on the IBM type 701, MTAC, Vol. 10, No. 53, 8-13.
168. P. C. JOHNSON and F. C. UFFELMAN, 1949. A punched card application of the Monte Carlo method, Proc. seminar on sci. comput., IBM, New York.
169. H. L. JONES, 1959. How many of a group of random numbers will be usable in selecting a particular sample, J. Amer. Statist. Assoc., Vol. 54, No. 285, 102-122.
170. H. JUDSON, Jr., G. G. SMITH and D. KLEIN, 1960. Experimental check of a Monte Carlo-calculated distribution of resonance neutron capture in a gold rod, Nucl. Sci. and Eng., Vol. 7, No. 3, 263-267.

171. M. L. JUNCOSA, 1954. Random number generation on the BRL high-speed computing machines. Math. reviews, Vol. 15, No. 6, 559.

172. H. KAHN, 1949. Modification of the Monte Carlo method, Proc. seminar on sci. comput., IBM, New York, 20-27.

173. H. KAHN, 1950. Random sampling (Monte Carlo) techniques in neutron attenuation problems, I, Nucleonics, Vol. 6, No. 5, 27-33.

174. H. KAHN, 1950. Random sampling (Monte Carlo) techniques in neutron attenuation problems, II, Nucleonics, Vol. 6, No. 6, 60-65.

175. H. KAHN, 1956. Use of different Monte Carlo sampling techniques, Symposium on Monte Carlo methods, 146-190.

176. H. KAHN and T. E. HARRIS, 1951. Estimation of particle transmission by random sampling, NBS Appl. Math. Series, No. 12, 27-30.

177. H. KAHN and A. W. MARSHALL, 1953. Methods of reducing sample size in Monte Carlo computations, J. Oper. Research Soc. Amer., Vol. 1, No. 5, 263-278.

178. M. H. KALOS, 1959. Importance sampling in Monte Carlo calculations of thick shield penetration, Nuclear Sci. and Eng., Vol. 2, Suppl., No. 1, 34-35.

179. M. H. KALOS and H. S. WILF, 1957. Monte Carlo solves reactor problems, Nucleonics, Vol. 15, No. 5, 64-68.

180. M. G. KENDALL and B. B. SMITH, 1938. Randomness and random sampling numbers, J. Royal. Statist. Soc., Vol. 101, No. 1, 147-166.

181. M. G. KENDALL and B. B. SMITH, 1939. Second paper on random sampling numbers, J. Royal. Statist. Soc., Vol. 6, Suppl., No. 1, 51-61.

182. M. G. KENDALL and B. B. SMITH, 1939. Random sampling numbers, Tracts for computers, Vol. 24, London.

183. T. J. KENNET and L. M. BOLLINGER, 1959. The multiple scattering of neutrons at resonances, Nuclear Phys., Vol. 12, 249-260.

184. G. W. KING, 1949. Stochastic methods in quantum mechanics, Proc. seminar on sci. comput., IBM, New York.

185. G. W. KING, 1949. Further remarks on stoachastic methods in quantum mechanics, Proc. seminar on sci. comput., IBM, New York, 92-94.

186. G. W. KING, 1951. Monte Carlo method for solving diffusion problems, Industr. Eng. Chemistry, Vol. 43, 2475.

187. G. W. KING, 1951. Stochastic methods in statistical mechanics, NBS Appl. Math. Series, No. 12, 12-18.

188. G. W. KING, 1953. The Monte Carlo method as a natural mode of expression in operation research, J. Oper. Research Soc. Amer., Vol. 1, 45-51.

189. D. H. LEHMER, 1949. Mathematical methods in large-scale computing units, Proc. sympos. on large-scale digital calcul. machinery, Harvard Univ. Press, 141-146.

190. E. J. LYTLE, 1956. A description of the generation and testing of a set of random normal deviates, Symposium on Monte Carlo methods, 234-248.

191. N. MANTEL, 1953. An extension of the Buffon needle problem, Ann. Math. Statistics, Vol. 24, 674-677.

192. W. C. MARCUS and L. NELSON, 1951. Methods of probabilities in chains applied to particle transmission through matter, NBS Appl. Math. Series, No. 12, 9-11.

193. A. W. MARSHALL, 1956. The use of multi-stage sampling schemes in Monte Carlo computations, Symposium on Monte Carlo methods, 123-140.

194. M. A. MARTINO, 1959. On the acceleration of Monte Carlo calculations, Nuclear Sci. and Eng., Vol. 2, Suppl, No. 1, 236-237.

195. W. MATTHES, 1960. Berechnung des Fermi-Alters und des Leckfaktors von heterogenen Leichtwasser-Reaktoren mit der Monte-Carlo-Methode, Nukleonik, Vol. 2, No. 1, 21-31.

196. M. MAYER, 1951. Report on a Monte Carlo calculation performed with the Eniac, NBS Appl. Math. Series, No. 12, 19-20.

197. McCRACKEN, 1955. The Monte Carlo method, Scient. American, Vol. 192, No. 5, 90.

198. N. METROPOLIS, 1956. Phase shifts—middle squares-wave equation, Symposium on Monte Carlo methods, 29-36.

199. N. METROPOLIS, R. BIVINS, M. STORM, A. TURKEVICH, J. M. MILLER and G. FRIEDLANDER, 1958. Monte Carlo calculations on intranulcear cascades, I, Low-energy studies, Phys. Rev., Vol. 110, 185-203.

200. N. METROPOLIS, R. BIVINS, M. STORM, J. M. MILLER, G. FRIEDLANDER and A. TURKEVICH, 1958. Monte Carlo calculations on intranuclear cascades, II, High-energy studies and pion processes, Phys. Rev., Vol. 110, 204-219.

201. N. METROPOLIS, G. REITWIESNER and J. NEUMANN, 1950. Statistical treatment of first 2000 decimal digits of e and π calculated on the Eniac, MTAC, Vol. 4, 109-111.

202. N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH and A. H. TELLER, 1953. Equation of state calculation by fast computing machines, J. Chem. Phys. Vol. 21, No. 6, 1087-1092.

203. N. METROPOLIS and S. ULAM, 1949. The Monte Carlo method, J. Amer. Statistical Assoc., Vol. 44, No. 247, 335-341.

204. N. MIYATAKE, T. MIKAMI, H. HIRAI and H. SUGIYAMA, 1958. Monte Carlo method of computer programming (Japanese), Sugaku, Vol. 9, No. 4, 238-240.

205. Monte Carlo method (proceedings of a symposium), 1951. NBS Appl. Math. Series, No. 12.

206. R. G. MOORE, 1953. A sequential test for randomness. Biometrika, Vol. 40, 111-115.

207. D. MORGENSTERN, 1955. Statistische Begründung numerischer Quadratur, Math. Nachrichten, Vol. 13, No. 3-4, 161-164.

208. K. W. MORTON, 1957. A generalization of the antithetic variate technique for evaluating integrals, J. Math. Phys., Vol. 36, No. 3, 289-293.

209. K. W. MORTON, 1956. On the treatment of Monte Carlo methods in text books, MTAC, Vol. 10, 223-224.

210. K. W. MORTON, 1957. Scaling neutron tracks in Monte Carlo shielding calculations, J. Nuclear Energy, Vol. 5, No. 3-4, 320-324.

211. J. MOSHMAN, 1954. The generation of pseudo-random numbers on a decimal calculator, J. Assoc. Comput. Machinery, Vol. 1, 88-91.

212. J. MOSHMAN, 1958. The application of sequential estimation to computer simulation and Monte Carlo procedures, J. Assoc. Comput. Machinery, Vol. 5, No. 4, 343-352.

213. M. MOTOO, 1959. Some evaluations for continuous Monte Carlo method by using Brownian hitting process, Ann. Inst. Statist. Math., Tokyo, Vol. 11, 49-54.

214. YAMASITA MOTOOKA, 1954. Methods of obtaining random numbers by means of digital computer (Japanese), J. Inst. Electr. Eng. Japan, Vol. 74, No. 5, 579-588.

215. T. S. MOTZKIN, 1956. Neighbor sets for random walks and difference equations, Symposium on Monte Carlo methods, 47-51.

216. M. E. MULLER, 1956. Some continuous Monte Carlo methods for the Dirichlet problem, Ann. Math. Statistics, Vol. 27, No. 3, 569-589.

217. M. E. MULLER, 1958. An inverse method for the generation of random normal deviates on large-scale computers, MTAC, Vol. 12, 167-174.

218. M. E. MULLER, 1959. A note on a method for generating points uniformly on N-dimensional spheres, Communs. Assoc. Comput. Mach., Vol. 2, No. 4, 19-20.

219. M. E. MULLER, 1959. A comparison of methods for generating normal deviates on digital computers, J. Assoc. Comput. Machinery, Vol. 6, No. 3, 376-383.
220. F. J. MUSK, 1959. A Monte Carlo simulation of a production planning problem, Comput. J., Vol. 2, No. 2, 90-94.
221. K. N. NAIR, 1938. On Tippett's random sampling numbers, Sankhya, Vol. 4, 65-72.
222. G. NEOVIUS, 1955. Artificial traffic trials using digital computers, Ericsson Techn., Vol. 11, No. 2, 279-291.
223. J. NEUMANN, 1951. Various techniques used in connection with random digits, NBS Appl. Math. Series, No. 12, 36-38.
224. A. OPLER, 1953. Application of computing machines to ion exchange column calculations, Industr. and Engineering Chemistry, Vol. 45, 2621-2633.
225. A. OPLER, 1951. Monte Carlo matrix calculations with punched card machines, MTAC, Vol. 5, No. 35, 115-120.
226. E. S. PAGE, 1954. The Monte Carlo solution of some integral equations, Proc. Cambridge Philos. Soc., Vol. 50, No. 3, 414-425.
227. E. S. PAGE, 1959. Pseudo-random elements for computers, Appl. Statist., Vol. 8, No. 2, 124-131.
228. Z. PAWLAK, 1956. Flip-flop as a generator of random binary digits, MTAC, Vol. 53, No. 10.
229. L. G. PECK, 1953. On uniform distribution of algebraic numbers, Proc. Amer. Math. Soc., Vol. 4.
230. J. F. PERKINS, 1955. Monte Carlo calculations of gamma-ray albedos of concrete and aluminium, J. Appl. Phys., Vol. 26, No. 6, 655.
231. J. F. PERKINS, 1955. Some Monte Carlo calculations of gamma-ray penetrations, J. Appl. Phys., Vol. 26, No. 11, 1372-1377.
232. I. G. PETROWSKY, 1934. Über das Irrfahrtproblem, Math. Ann., Vol. 109, 425.
233. M. H. QUENOUILLE, 1959. Tables of random observations from standard distributions, Biometrika, Vol. 46, No. 1-2, 178-202.
234. RAND Corporation, 1955. A Million Random Digits With 100,000 Normal Deviates. The Free Press.
235. R. D. RICHTMYER, R. VAN NORTON and A. WOLFE, 1958. The Monte Carlo calculation of resonance capture in reactor lattices, Proceedings of the Second International Geneva Conference on the Peaceful Uses of Atomic Energy, No. 2489.
236. M. ROSENBLUTH and A. ROSENBLUTH, 1954. Further results on Monte Carlo equations of state, J. Chemical Phys., Vol. 22, No. 5, 881.

237. A. ROTENBERG, 1960. A new pseudo-random number generator, J. Assoc. Comput. Machinery, Vol. 7, No. 1, 75-77.
238. A. ROTENBERG, A. LAPIDUS and E. WETHERELL, 1959. A Monte Carlo calculation of thermal utilization, Nuclear Sci. and Eng., Vol. 6, No. 4, 288-293.
239. W. ROTHENSTEIN, 1960. Some Monte Carlo and analytical results for resonance capture in lattices, Nuclear Sci. and Eng., Vol. 8, 122-127.
240. F. SCHEID, 1958. Radial distribution of the center of gravity of random points on a unit circle, J. Research NBS, Vol. 60, No. 4, 307-308.
241. D. O. SCHNEIDER and D. V. CORMACH, 1959. Monte-Carlo calculations of electron energy loss, Radiation Research, II, 418-429.
242. B. A. SHOOR, L. NELSON, W. MARCUS and R. L. ECHOLS, 1951. A Monte Carlo technique for estimating particle attenuation in bulk matter, NBS Appl. Math. Series, No. 12, 24-26.
243. G. SPENCER, 1955. Random numbers and their generation, Computers and Automation, Vol. 4, No. 3, 10-11.
244. B. J. SPINARD, G. H. GOERTZEL and W. S. SNYDER, 1951. An alignment chart for Monte Carlo solution of the transport problem, NBS Appl. Math. Series, No. 12, 4-5.
245. F. STERZER, 1959. Random number generator using subharmonic oscillators, Rev. Scient. Instrum., Vol. 30, No. 4, 241-243.
246. H. SUGIYAMA and O. MIYATAKE, 1959. Design of "random walker" for Monte Carlo method, I, Theory, J. Inst. Polytech., Osaka City Univ., ser A, Vol. 10, 35-41.
247. Symposium on Monte Carlo methods, 1956. Ed. H. A. Meyer, Wiley, New York.
248. O. TAUSSKY and J. TODD, 1956. Generation and testing of pseudorandom numbers, Symposium on Monte Carlo methods, 15-28.
249. L. H. THOMAS, 1953. A comparison of stochastic and direct methods for the solution of some special problems, J. Operat. Research, Soc. America, I, 181-186.
250. L. H. C. TIPPETT, 1927. Random sampling numbers, Tracts for computers, Vol. 15, London.
251. K. D. TOCHER, 1954. The application of automatic computers to sampling experiments, J. Royal Statistical Soc., ser. B, Vol. 16, No. 1, 39-61.
252. J. TODD, 1953. Experiments on the inversion of a 16 × 16 matrix, NBS Appl. Math. Series, No. 29, 113-115.
253. J. TODD, 1954. Experiments in the solution of differential equations by Monte Carlo methods, J. Washington Acad. Sci., Vol. 44, No. 12, 377-381.

254. H. F. TROTTER and J. W. TUKEY, 1956. Conditional Monte Carlo for normal samples, Symposium on Monte Carlo methods, 64-79.

255. J. W. TUKEY, 1957. Antithesis or regression? Proc. Cambridge Philos. Soc., Vol. 53, No. 4, 923-924.

256. S. ULAM, 1949. On the Monte Carlo method, Proc. of a Second Symposium on large scale digital calculating machinery, Harvard Univ. Press, 207-212.

257. S. ULAM, 1950. Random processes and transformations, Proc. Internat. Congress of Math., 264-275.

258. S. ULAM, 1956. Application of Monte Carlo methods to tactical games, Symposium on Monte Carlo methods, 63.

259. C. W. VICKERY, 1939. On drawing a random sample from a set of punched cards, J. Royal Statistical Soc., Suppl., Vol. 6, 62-66.

260. D. F. VOTAW and J. A. RAFFERTY, 1951. High speed sampling, MTAC, Vol. 5, No. 33, 1-8.

261. J. E. WALSH, 1949. Concerning compound randomization in the binary system, Ann. Math. Statistics, Vol. 20, 580-589.

262. J. E. WALSH, 1956. Questionable usefulness of variance for measuring estimate accuracy in Monte-Carlo importance sampling problems, Symposium on Monte Carlo methods, 141-144.

263. J. E. WALSH, 1956. A Monte Carlo technique for obtaining tests and confidence intervals for insurance mortality rates, Symposium on Monte Carlo methods, 265-277.

264. A. WALTHER, 1956. Experiments and models for the Monte Carlo method, Symposium on Monte Carlo methods, 278-282.

265. W. WASOW, 1956. A note on the inversion of matrices by random walks, MTAC, Vol. 5, No. 38, 78-81.

266. W. WASOW, 1951. Random walks and the eigenvalues of elliptic difference equations, J. Research NBS, Vol. 46, 65-73.

267. W. WASOW, 1951. On the mean duration of random walks, J. Research, NBS, Vol. 46, 462-472.

268. W. WASOW, 1951. On the duration of random walks, Ann. Math. Statistics, Vol. 22, 199-216.

269. W. WASOW, 1953. Metodi probabilistici per la risoluzione numerica di alcuni problemi di analisi, Atti 4 congr. Unione mat. ital. Vol. 2, 248-250.

270. W. WASOW, 1953. Metodi probabilistici per la risoluzione numerica di alcuni problemi di analisi e di algebra, Rend. mat. e appl. Univ. Roma, Vol. 11, 336-346.

271. J. G. WENDEL, 1957. Groups and conditional Monte Carlo, Ann. Math. Statistics, Vol. 28, 1048-1052.

272. R. R. WILSON, 1951. The range and straggling of high energy electrons, Phys. Rev., Vol. 84, No. 1, 100-103.

273. R. R. WILSON, 1951. Showers produced by low-energy electrons and photons, NBS Appl. Math. Series, No. 12, 1-3.

274. R. R. WILSON, 1952. Monte Carlo study of shower production, Phys. Rev., Vol. 86, No. 3, 261-269.

275. W. W. WOOD and J. D. JACOBSON, 1957. Preliminary results from a recalculation of the Monte Carlo equation of state of hard spheres, J. Chem. Phys., Vol. 27, 1207-1208.

276. W. W. WOOD and F. R. PARKER, 1957. Monte Carlo equation of state of molecules interacting with the Lennard-Jones potential, I, Supercritical isotherm at about twice the critical temperature, J. Chem. Phys., Vol. 27, 720-733.

277. W. W. WOOD, F. R. PARKER and J. D. JACOBSON, 1958. Recent Monte Carlo calculations of the equation of state of Lennard-Jones and hard sphere molecules, Nuovo cimento, Vol. 9, Suppl., No. 1, 133-143.

278. W. W. WOODBURY, 1949. Monte Carlo calculations, Proc. seminar on sci. comput., IBM, New York.

279. L. C. YOUNG, 1941. On randomness in ordered sequences, Ann. Math. Statistics, Vol. 12, 293-300.

280. E. C. YOWELL, 1949. A Monte Carlo method of solving Laplace's equation, Proc. seminar on sci. comput., IBM, New York.

281. G. U. YULE, 1938. A test of Tippett's random sampling numbers, J. Royal Statistical Soc., Vol. 101, No. 1, 167-172.

282. N. ZIERLER, 1959. Linear recurring sequences, J. Soc. Industr. and Appl. Math., Vol. 7, No. 1, 31-48.

Index

Absorbing state, 31
Accuracy of Monte Carlo method, 28, 88
Algebraic linear systems, 19
Analysis of showers, 86
Angle, azimuthal (scattering), 94, 95
Anisotropic scattering, 95
Aperiodicity stretch, 204, 209
—, length of, 209
Approximate construction of Brownian trajectories, 74
Approximate evaluation of integrals, 65
Area, computation of, 40

Bakhvalov, N. S., 70
Barn, 92
Berger, M., 113
Bernoulli theorem (law of large numbers), 7
Boundary value problems, solution of, 23
Brownian trajectories, 4, 74
—, approximate construction, 74
Busy time of line, 135

Cascade (shower), 85, 86
Chain, Markov, 31, 250
Chavchanidze, V. V., 223
Chebyshev inequality, 8, 9
Coding, 194
Collision density, 112
Continuous integral, 73
Convergence, increased rate of, 68
—, rate of (of Monte Carlo method), 46
Coordinate system, center-of-mass, 94
—, laboratory, 94
Correlation coefficient, 57, 242
Covariance function, 241
Criterion, chi-square, 214
—, Kolmogorov, 215
—, omega-square, 218
—, sequence, 213
Critical values of reactor parameters, 124
Critical volume, 118
Cross section, capture, 92
—, differential (for elastic scattering), 95
—, effective neutron, 92
—, fission, 92
—, interaction, 85
—, partial, 85
—, scattering, 92
—, transverse (of absorption, total), 92

Danil'chenko, I. A., 230
Decoding of information, 194
Detection curves, 174
—, of signal probability, 173, 184
Deterministic quasi-random points, 83
Differential elastic scattering cross section, 95
Digits, random, 279
Dirichlet problem, 24

Distribution, normal (Gaussian), 14, 259
—, —, multidimensional, 266
—, Pareto, 194, 256
—, Poisson, 256, 262
—, quasi-uniform, 245
—, Rayleigh, 165, 255
—, Rice, 165, 261, 271
—, uniform, 244
Duration of random walk, 27, 34

Effective neutron cross sections, 92
Effective neutron multiplication coefficient, 118, 120
Elastic scattering, 94, 95, 103, 105
Elliptic equation, 28
Empirical entropy, 240
Ergodic Markov chains, 34
Erlangian stream, 142
Error, Monte Carlo method, 8, 41, 88
—, probability, 42, 88
—, standard, 42

False-alarm probability, 173, 184
Fibonacci series, 200
Fictitious mass (statistical weight), 89, 103, 106, 110, 123
Fraction of losses, 153
—, mean, 155
Free path, 99
—, length of, 99, 102

Gas-discharge tube, 221
Gauss' theorem, 10, 14
Generalized Halton sequence, 82
Generation of random numbers, 13
Generations, 87, 119
—, determining number of, 119
Generator, noise, 220
Group sampling, 50

Halton's sequence, 80
—, generalized, 82
Hypersphere, volume of, 219

Inclusion of singularity in density, 63
Inelastic scattering, 96, 104
Infinitely dimensional integrals, 82
Instantaneous stream density, 133
Integrals, infinite-dimensional, 82
—, multidimensional, 61, 79
—, Wiener (continuous), 73
Isotropic scattering, 94, 96

Kan, G., 113
Kellogg, O., 126
Kendall, M. G., 210

Laboratory system of coordinates, 94
Laplace equation, 24, 29, 35
Law of distribution of the fraction of losses, 153
Law of large numbers (Bernoulli theorem), 7
Lehmer, D., 199, 201
Length of aperiodicity stretch, 209
Length of free path, 99, 102
Lethargy, 115
Lexicographic analysis of a shower, 87
Lifetime of system, 32, 34
Likelihood ratio, 172

Linear algebraic systems, 19
Locally random set of numbers, 210
Lyapunov, A. M., 14

Macroscopic cross section, 92
Markov chain, 31, 250
—, ergodic, 34
—, terminating, 31
Maximum—likelihood method, 120, 190
"Mean-product" method, 199
Microscopic cross section, 92
"Mid-square" method, 199
Mixing method, 202
Moments, method of, 121
Monte Carlo method, efficiency of, 45
Morton, K.,110
Multidimensional integrals, evaluation of, 61, 79

Neumann, J. von, 199
Neumann problem, 35
Neutron fission, 103
Neutron multiplication coefficient, 118, 120
Noise, 157
Noise generator, 220
Normal (Gaussian) distribution, 259
Normal multidimensional distribution, 266
Normal quantities, 279
Normalization of number of neutrons, 122
Numbers, pseudo-random, 18, 83, 196, 279

Palm's formula, 132
Pareto distribution, 194, 256
Partial cross sections, 85
Pawlak, Z., 223
Pearson, K., 170
Peierls equation, 125
Periodicity of a pseudo-random sequence, 208
"Perturbation" method, 208
Photoelectric effect, 85
Photon, 85
Piecewise approximation method, 262
Poisson's law, 16, 132, 141
—, simulation of, 256, 262
Principal part, separation of, 46
Probability, false-alarm, 173, 184
—, reflection, 108, 114
—, signal-detection, 173, 184
—, transition, 250
—, transmission, 106, 109, 114
Probability density, amplitudes, 163, 167
—, noise, 161
Probability distortion, 206
Probable error, 42, 106
Pseudo-random numbers, 196, 279
—, for BESM computer, 204
—, for ENIAC, 199
—, for OARAC, 200
—, for SEAC, 201
—, for "Strela", 202
—, for SVAC, 200
—, for UNIVAC, 200
—, for "Ural-1", 205

Quasi-random points in a Monte Carlo system, 78
Quasi-random sequence of numbers, 18
Quasi-uniform distribution, 245

Radar detection of target, 188
Radioactive random—number generators, 230

Radio-relay system, 192
Random digits, 279
Random direction, 98
Random events, 13
Random-number generator, for BESM computer, 225
—, for G-2, 231
—, for "Strela", 226, 230, 238
Random-number generators, 2, 16, 197, 220
Random numbers, generation of, 13
"Random" quadrature formulas, 57
Random streams of calls, generation of, 138
Random walk in a lattice, 24, 33
Reactor, 118
Region, critical, 118
—, subcritical, 118
—, supercritical, 118
Runs, 212

Scattering, elastic, anisotropic, 95, 103
—, —, isotropic, 94, 105
—, inelastic, 96, 103
Selection of random direction, 98
Selective sampling, 48, 90
Separation of principal part, 46
Sequence, Halton's, 80
Sequential analysis, 179, 187
"Shot noise", 221
Signal, 156
—, quantized, 169
Similar trajectories, method of, 110
Simulation—of distribution of fission neutrons, 119
—, of free path, 100
—, of Markov chains, 33, 250
—, of a physical process, 85, 104, 119
—, of random events, 13
—, of random vectors, 264
—, of scattering density, 112
—, of (a system of) random events, 247
Smiryagin, V. P., 221
Smith, B. B., 210
Solution—of boundary value problems, 23
—, of systems of linear algebraic equations, 4, 19
—, of transport equations, 115
Spectrum, neutron, 114
—, —, fission, 98, 119
Square-law processing of noise-like signals, 176, 184
State, absorbing, 31
—, connected, 31
Stationary random process, 240
Stationary stream, 130, 141
Statistical estimate of error, 8, 41, 88
Statistical weight, fictitious mass, 89, 103, 106, 110, 123
—, replacing Monte Carlo determination of interactions, 103
—, representing absorption, 106
Sterzer, F., 228
Stream, Erlangian, 142
—, nonordinary, 133, 147
—, of calls, 127
—, of homogeneous events, 130, 134, 138
—, ordinary, 130, 139, 141
—, simple, 131, 139
—, stationary, 130, 141
—, with fixed minimum interval, 144
—, with limited after—effect, 130, 139, 141
—, with substantial inhomogeneity, 153
—, with variable parameter, 146
—, with uniform distribution of intervals, 141

Stream density, 131, 143
Subcritical volume (region), 118
Subharmonic generator, 228
Supercritical volume (region), 118
Symmetrization, 53
System, mass-service, 134
—, —, mixed, 135
—, —, with delay, 135
—, —, with rejection, 135

Table of normal quantities, 276, 279
Table of random numbers, 197, 273, 279
Target azimuth, 189
Terminating Markov chain, 31
Test of hypotheses, 171
Thermal conductivity equation, 29
Threshold signal, 174
Transition matrix, 37, 251
Transport (kinetic) equation, solution of, 115
Trigger, 223

Uniform (cubic) lattice, 70
Uniting of states, 37

Van Weingarden, 201
Valdimirov, V. S., 126
Volume, critical, 118
—, subcritical, 118
—, supercritical, 118
Volume of hypersphere, 219

Wald, A., 170